

Springer Protocols

Dilip Kumar Arora
Surajit Das
Mesapogu Sukumar
Editors

Analyzing Microbes

Manual of Molecular Biology
Techniques

 Springer

Springer Protocols

Dilip Kumar Arora
Surajit Das
Mesapogu Sukumar
Editors

Analyzing Microbes

Manual of Molecular Biology
Techniques



Springer

Editors

Dilip Kumar Arora
National Bureau of Agriculturally
Important Microorganisms
Maunath Bhanjan
Uttar Pradesh
India

Surajit Das
Department of Life Science
National Institute of Technology
Rourkela
Odisha
India

Mesapogu Sukumar
National Bureau of Agriculturally
Important Microorganisms
Maunath Bhanjan
Uttar Pradesh
India

ISSN 1949-2448 ISSN 1949-2456 (electronic)
ISBN 978-3-642-34409-1 ISBN 978-3-642-34410-7 (eBook)
DOI 10.1007/978-3-642-34410-7
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013931210

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The discipline microbiology is researched actively, and the field is advancing continually. It is estimated that only about 1 % of all of the microbe species on earth have been studied. Rapid advances in molecular biology have revolutionized the study of microorganisms in the environment and improved our understanding of the composition, phylogeny, and physiology of microbial communities. The advent of molecular biology has offered a number of revolutionary new insights into the detection and enumeration of soilborne microorganisms. DNA sequences provide information on identifying unknown species from 16S and ITS rRNA sequences of individual bacterial and fungal species. Molecular methods monitor both pathogens and also beneficial organisms in soils for detection and quantification. The in-depth exploitation of PCR potential led to more sophisticated variants of the technique (improved even from the currently expanding real-time PCR) that increases the speed and sensitivity in microbial identification and diagnostics. These molecular techniques provide new insights about their functions and interactions within ecological niches.

Analyzing Microbes—Manual of Molecular Biology Techniques is a practical guide to the application of important molecular biology techniques in microbiological research. The chapters are written by a group of international scientists who are recognized authorities in their research areas from universities/researchers and often the new techniques that are described. These volumes are aimed for graduate, postgraduate, Ph.D. students, and laboratory technicians working in different biotechnology/microbiology laboratories. It is also valuable to the larger community of researchers who have recognized the potential of genomics research and may be beginning to explore the technologies involved. Moreover, the volumes are also targeted as handouts for students, teachers, and researchers world over.

The central parts of the chapters are the experimental protocols which are presented so as to be readily used at the laboratory bench. Although a number of the procedures described represent the tried and trusted, we have striven to include variants on existing technologies that an experiment can be performed. These step-by-step protocols are intended to be concise and easy to follow. Suggestions to successfully apply the procedures are included, along with recommended materials and suppliers. A special feature of the chapters is that, in addition to the protocols, important background information and representative results of applying the methods are given. References are provided to enable the investigator to become better acquainted with

the topic. Researchers in any field that utilizes microbial systems will find this work of value. In addition to microbiology and bacteriology, this book highlights the current state-of-the-art molecular microbiology techniques in biotechnology, microbiology research, and environmental microbiology.

The aim of the book *Analyzing Microbes—Manual of Molecular Biology Techniques* has been to produce a self-contained laboratory manual which will be useful to both experienced practitioners and beginners in the field. We hope that this book stimulates your creativity and wish you success in your experiments.

Maunath Bhanjan, Uttar Pradesh, India
Rourkela, Odisha, India
Maunath Bhanjan, Uttar Pradesh, India

Dilip Kumar Arora
Surajit Das
Mesapogu Sukumar

Contents

1	Microbial DNA Extraction, Purification, and Quantitation . . .	1
	Sukumar Mesapogu, Chandra Mouleswararao Jillepalli, and Dilip K. Arora	
2	Fluorescent-Based Detection, Quantitation, and Expression of Viral Gene by qRT-PCR	17
	Shelly Praveen and Vikas Koundal	
3	Restriction Enzymes and Their Role in Microbiology	31
	Sukumar Mesapogu, Chandra Mouleswararao Jillepalli, and Dilip K. Arora	
4	Genetic Fingerprinting Techniques for Molecular Characterisation of Microbes	37
	Annette Reineke and K. Uma Devi	
5	Agarose Gel Electrophoresis and Polyacrylamide Gel Electrophoresis: Methods and Principles	73
	Sukumar Mesapogu, Chandra Mouleswararao Jillepalli, and Dilip K. Arora	
6	Molecular Identification of Microbes: I. <i>Macrophomina phaseolina</i>	93
	Bandamaravuri Kishore Babu, T. Kiran Babu, and Rajan Sharma	
7	Molecular Identification of Microbes: II. <i>Bacillus</i>	99
	Anil Kumar Saxena	
8	Molecular Identification of Microbes: III. <i>Pseudomonas</i>	105
	Bhim Pratap Singh and Ratul Saikia	
9	Molecular Identification of Microbes: IV. <i>Vibrio</i>	113
	Hirak Ranjan Dash, Neelam Mangwani, and Surajit Das	
10	Molecular Identification of Microbes: V. <i>methylophs</i>	123
	Kamlesh K. Meena, Manish Kumar, and D.P. Singh	
11	Preservation and Maintenance of Microbial Cultures	135
	Sudheer Kumar, Prem L. Kashyap, Ruchi Singh, and Alok K. Srivastava	

12	Microbes from Extreme Environment: Molecular Identification Procedures	153
	Surajit Das	
13	ELISA-Based Identification and Detection of Microbes	169
	Jyoti Verma, Sangeeta Saxena, and Sunil G. Babu	
14	Analysis of Microbial Diversity and Construction of Metagenomic Library	187
	Thangamani Rajesh, Jeyaprakash Rajendhran, and Paramasamy Gunasekaran	
15	Bioinformatics Tools for Interpretation of Data Used in Molecular Identification	209
	Suchi Smita, Krishna P. Singh, Bashir A. Akhoon, Shishir K. Gupta, and Shailendra K. Gupta	
16	Molecular Phylogenetics of Microbes	245
	Surajit Das and Hirak Ranjan Dash	
17	Microarray Technology: Basic Concept, Protocols, and Applications	261
	P.P. Dubey and Dinesh Kumar	
18	Microarray Analysis of Different Functional Genes of Microorganisms	281
	Hirak Ranjan Dash and Surajit Das	
19	DNA Cloning and Sequencing	291
	Bandamaravuri Kishore Babu, Anu Sharma, and Hari Kishan Sudini	
20	Biological Sequence Analysis: Algorithms and Statistical Methods	303
	Suchi Smita, Krishna P. Singh, Bashir A. Akhoon, and Shailendra K. Gupta	
	Appendix 1: Biosafety, GLP, and Biosecurity	335
	Appendix 2: Address for Instruments and Chemicals Suppliers . . .	339
	Index	345

Contributors

Bashir A. Akhoo CSIR-Indian Institute of Toxicology Research, Lucknow, India

Dilip K. Arora National Bureau of Agriculturally Important Microorganisms, Kusumaur, Maunath Bhanjan, Uttar Pradesh, India

Sunil G. Babu Department of Biotechnology, School of Bioscience and Biotechnology, Babasaheb Bhimrao Ambedkar University, Lucknow, Uttar Pradesh, India

Bandamaravuri Kishore Babu Cereals Pathology, International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Andhra Pradesh, India

Surajit Das Department of Life Science, National Institute of Technology, Rourkela, Odisha, India

Hirak Ranjan Dash Department of Life Science, National Institute of Technology, Rourkela, Odisha, India

P.P. Dubey Department of Animal Genetics and Breeding, Guru Angad Dev University of Veterinary, Animal Sciences, Ludhiana, Punjab, India

Paramasamy Gunasekaran Department of Genetics, Centre for Excellence in Genomic Sciences, School of Biological Sciences, Madurai Kamaraj University, Madurai, Tamil Nadu, India

Shishir K. Gupta Society for Biological Research and Rural Development, Lucknow, India

Shailendra K. Gupta CSIR-Indian Institute of Toxicology Research, Lucknow, India

Chandra Mouleswararao Jillepalli Department of Chemistry, Kakatiya University, Warangal, Andhra Pradesh, India

Prem L. Kashyap National Bureau of Agriculturally Important Microorganisms, Kusumaur, Maunath Bhanjan, Uttar Pradesh, India

T. Kiran Babu Cereals Pathology, International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Andhra Pradesh, India

Vikas Koundal Division of Plant Pathology, Advanced Center for Plant Virology, Indian Agricultural Research Institute, New Delhi, India

Manish Kumar National Bureau of Agriculturally Important Microorganisms, Maunath Bhanjan, Uttar Pradesh, India

Sudheer Kumar National Bureau of Agriculturally Important Microorganisms, Kusumaur, Maunath Bhanjan, Uttar Pradesh, India

Dinesh Kumar National Bureau of Animal Genetic Resources (NBAGR), Karnal, Haryana, India

Neelam Mangwani Department of Life Science, National Institute of Technology, Rourkela, Odisha, India

Kamlesh K. Meena National Bureau of Agriculturally Important Microorganisms, Maunath Bhanjan, Uttar Pradesh, India

Sukumar Mesapogu National Bureau of Agriculturally Important Microorganisms, Kusumaur, Maunath Bhanjan, Uttar Pradesh, India

Shelly Praveen Division of Plant Pathology, Advanced Center for Plant Virology, Indian Agricultural Research Institute, New Delhi, India

Jeyaprakash Rajendhran Department of Genetics, Centre for Excellence in Genomic Sciences, School of Biological Sciences, Madurai Kamaraj University, Madurai, Tamil Nadu, India

Thangamani Rajesh Department of Genetics, Centre for Excellence in Genomic Sciences, School of Biological Sciences, Madurai Kamaraj University, Madurai, Tamil Nadu, India

Annette Reineke Department of Phytomedicine, Geisenheim Research Centre, Geisenheim, Germany

Ratul Saikia Biotechnology Division, CSIR-North East Institute of Science and Technology, Jorhat, Assam, India

Sangeeta Saxena Department of Biotechnology, School of Bioscience and Biotechnology, Babasaheb Bhimrao Ambedkar University, Lucknow, Uttar Pradesh, India

Anil Kumar Saxena Division of Microbiology, Indian Agriculture Research Institute, New Delhi, India

Anu Sharma National Bureau of Agriculturally Important Microorganisms, Kusumaur, Maunath Bhanjan, Uttar Pradesh, India

Rajan Sharma Cereals Pathology, International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Andhra Pradesh, India

D.P. Singh National Bureau of Agriculturally Important Microorganisms, Maunath Bhanjan, Uttar Pradesh, India

Ruchi Singh National Bureau of Agriculturally Important Microorganisms, Kusumaur, Maunath Bhanjan, Uttar Pradesh, India

Krishna P. Singh CSIR-Indian Institute of Toxicology Research, Lucknow, India

Bhim Pratap Singh Department of Biotechnology, Mizoram University, Mizoram, India

Suchi Smita Society for Biological Research and Rural Development, Lucknow, India

Alok K. Srivastava National Bureau of Agriculturally Important Microorganisms, Kusumaur, Maunath Bhanjan, Uttar Pradesh, India

Hari Kishan Sudini International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India

K. Uma Devi Department of Botany, Andhra University, Visakhapatnam, India

Jyoti Verma Department of Biotechnology, School of Bioscience and Biotechnology, Babasaheb Bhimrao Ambedkar University, Lucknow, Uttar Pradesh, India

Chapter 1

Microbial DNA Extraction, Purification, and Quantitation

**Sukumar Mesapogu, Chandra Mouleswararao Jillepalli,
and Dilip K. Arora**

Abstract

Cell wall of microorganisms is broken by chemical or enzymatic lysis or a combination of both. Generally lysozyme is used to digest the rigid cell wall structure which has high amounts of lipid while detergent like SDS solubilizes the phospholipids in the cell membrane. EDTA destabilizes the cell envelope and deactivates the DNases by chelating with the magnesium ions in the membranes which are essential for integrity of cell envelope. Insoluble cell debris is removed via centrifugation, leaving an upper aqueous suspension containing the DNA, proteins, and RNA. Purification of DNA from proteins can be achieved by various methods, generally by protease treatment, to hydrolyze the proteins resulting in water-soluble amino acids or shaking the aqueous suspension with phenol chloroform. The aqueous phenol emulsion is then separated by centrifugation. Proteins (having both hydrophobic as well as hydrophilic amino acid residues) get collected at the interphase. While RNA can be removed by RNase treatment, DNA can be concentrated by addition of ice-chilled ethanol or isopropanol and precipitated DNA is collected as pellet by centrifugation. This chapter describes the protocol to check the purity and quantify DNA.

1.1 Introduction

Isolation of genomic DNA from microorganisms has become a useful tool to determine the fates of selected microorganisms or recombinant genes and to reveal genotypic diversity and its change in microbial ecosystems. The protocols in this chapter provides a frame work for isolating high quality genomic DNA from a variety of organisms, including bacteria, plasmid DNA [1], actinomycetes, yeast [2, 3], and fungi [4]. All of these protocols yield high molecular weight (HMW) DNA, which remains of high quality (i.e., not degraded in to smaller fragments) for several years when stored as specified below. For each organism a specific procedure is provided for releasing free chromosomal DNA from its cellular or nuclear location. The first task in each of these protocols is the removal of cell wall that is typically lysed in an SDS solution containing sucrose. The released DNA is prevented

from degrading DNAses and other proteins by EDTA and proteinase-K respectively the cellular proteins [2].

The chromosomal DNA of the *Escherichia coli* is a large circular molecule of approximately 3.2 kb size. The DNA is attached to the plasma membrane at many points. Being large in size, DNA is prone to mechanical breakage. However, if extraction is performed carefully, large fragments of chromosomal DNA can be obtained with an average length of 1–2 kb. The bacterial cell wall is enclosed in a cytoplasmic membrane and surrounded by a rigid cell wall. With some species, including *E. coli*, the cell wall may itself be enveloped by second outer membrane. All of these barriers have to be disrupted to release the cell components. Techniques for breaking open bacterial cells can be divided into physical methods, in which cells are disrupted by mechanical forces and chemical methods, where cell lysis is brought about by exposure to chemical agents that affect the integrity of the cell barriers. Chemical methods are most commonly used with bacterial cells when the object is DNA preparation. Chemical lysis generally involves one agent attacking the cell wall and another disrupting the cell membrane. The chemicals that are used depend on the species of bacterium involved, but with *E. coli* and related organisms, weakening of the cell wall is usually brought about by lysozyme, ethylenediamine tetraacetate (EDTA), or combination of both. Lysozyme is an enzyme that is present in egg white and in secretions such as tears and saliva, and which digests the polymeric compounds that give the cell wall its rigidity. On the other hand, EDTA removes magnesium ions that are essential for preserving the overall structure of the cell envelope, and also inhibits cellular enzymes that could degrade DNA. Under some conditions, weakening the cell wall with lysozyme or EDTA is sufficient to cause bacterial cells to burst, but usually a detergent such as sodium dodecyl sulfate (SDS) is also added [5]. Detergents aid the process of lysis by removing lipid molecules and thereby cause disruption of the cell membranes. Having lysed the cells, the final step in preparation of a cell extract is removal of insoluble cell debris. Component such as partially digested cell wall fractions can be pelleted by centrifugation, leaving the cell extract as a reasonably clear supernatant. Most protocols for the preparation of genomic DNA consist of lysis, followed by incubation with a nonspecific protease and a series of extractions prior to precipitation of the nucleic acids. Such procedures effectively remove contaminating proteins, but are not effective in removing exopolysaccharides which can interfere with the activity of enzymes such as restriction endonucleases and ligases. In this unit, however, the protease incubation is followed by a CTAB extraction whereby CTAB complexes with both polysaccharides and residual protein, effectively removing both in the subsequent emulsification and extraction. This procedure is effective in producing digestible chromosomal DNA from a variety of gram-negative bacteria, all

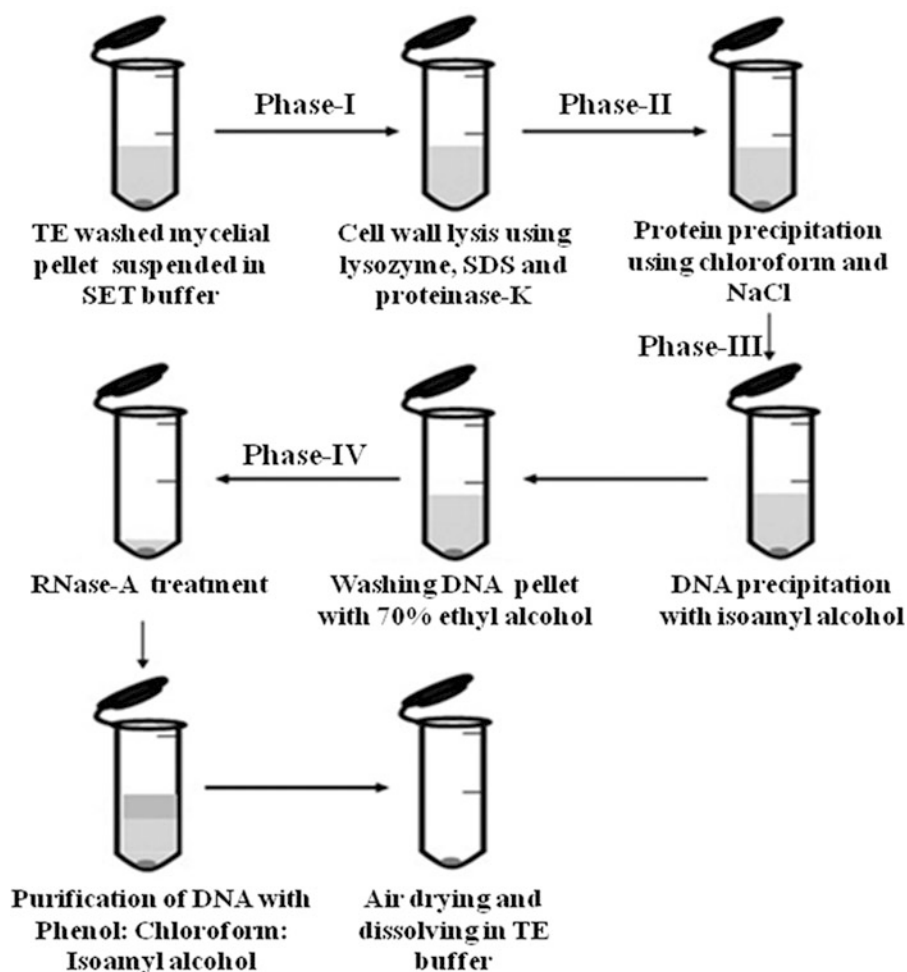


Fig. 1.1 General steps involved in genomic DNA isolation (Phase I: Cell lysis, Phase II: Protein degradation and precipitation, Phase III: DNA precipitation, Phase IV: RNA degradation and precipitation)

of which normally produce large amounts of polysaccharides. The actinomycetes are gram-positive bacteria which have a characteristically high G + C content in their DNA (>55 %). Many species produce a wide variety of secondary metabolites, including anti-helminthic compounds, antitumor agents, and the majority of known antibiotics, which have been exploited by their use in medicine and agriculture. The actinomycetes were originally considered to be an intermediate group between bacteria and fungi but now recognized as prokaryotic. The detailed protocol is represented as four phases as shown in Fig. 1.1

It is based on the conformational difference between plasmid and chromosomal DNA of the bacteria. Plasmid molecules are double stranded, circular entities, and generally exist as covalently closed circular molecule (supercoiled form), while chromosomal DNA molecule exists as linear double stranded molecule. So use is

made of the fact that linear double stranded DNA is denatured by exposing to high pH values of the lysing solution (in the range of pH 12–12.5). On the other hand, the covalently closed circular plasmid DNA in supercoiled form is resistant to these conditions. So when the pH is lowered in further steps, renaturation takes place and plasmid renatures faster (if denatured) than the chromosomal DNA. Because the renaturation is done at cool temperature and also pH lowering is sharp, as a result the whole chromosomal DNA forms an insoluble clump (aggregate) because of the mis-matched base-pairing. The aggregate DNA can easily be separated by centrifugation, as plasmid remains in the supernatant while aggregate chromosomal DNA forms pellet [5].

The need to adapt organic extraction methods to take account of the biochemical contents of different types of starting material has stimulated the search for DNA purifications methods that can be used with any species. This is one of the reasons why ion-exchange chromatography has become so popular. A similar method involves a compound called guanidinium thiocyanate, which has two properties that make it useful for DNA purification. First it denatures and dissolves all biochemicals other than nucleic acids and can therefore be used to release DNA virtually from any type of cell or tissue. Second, guanidinium thiocyanate allows DNA to bind tightly to silica particles. This provides an easy way of recovering the DNA from the denatured cell extracts [6]. One possibility is to add the silica directly to the cell extract but, as with the ion-exchange methods, it is more convenient. In addition to DNA, the cell extract contain significant quantities of protein and RNA. A variety of methods can be used to purify the DNA from this mixture. One approach is to treat mixture with reagents which degrade the contaminants, leaving a pure solution of DNA. The standard way to deproteinize a cell extract is to add phenol or a 1:1 mixture of phenol and chloroform. These organic solutions precipitate proteins but leave the nucleic acids (DNA and RNA) in an aqueous solution. The result is that if the cell extract is mixed gently with the solvent and the layers then separated by centrifugation, precipitated protein molecules are left as a white coagulated mass at the interface between the aqueous and organic layers. The aqueous solution of nucleic acids can then be removed with a pipette [7].

With some cell extract, the protein content is so great that a single phenol extraction is not sufficient to purify nucleic acids completely. This problem could be solved by carrying out several phenol extractions one after the other, but this is undesirable as each mixing and centrifugation step results in a certain amount of breakage of the DNA molecules. This can be solved by treating the cell extract with protease such as pronase or protease K before phenol extraction. These enzymes break polypeptides down into smaller units, which are more easily removed by phenol. Some

RNA molecules, especially mRNA, are removed by phenol treatment, but most remain with the DNA in the aqueous layer. The only effective way to remove the RNA is with the enzyme ribonuclease, which rapidly degrades these molecules into ribonucleotide subunits. A second useful method is drop dialysis, which can remove salt, SDS, and even some enzyme inhibitors. As such, it can be used with many methods involving DNA purification before or after enzymatic reactions. DNA fragments larger than a few 100 base pairs can be separated from smaller fragments by chromatography on a size exclusion column such as Sephacryl S-500. To simplify this procedure, the following minispin column method has been developed [6]. For fragments from 200 bp to 10 kb the agarose purification is ideal. For smaller fragments (20–400 bp), the acrylamide purification is preferred.

Ultra violet (UV) spectrophotometry is most commonly used for the determination of DNA concentration. The resonance structure of pyrimidine and purines are responsible for these absorptions. The DNA has a maximum and minimum absorbance at 260 nm. However, these are strongly affected by the degree of base ionization and hence pH of the measuring medium. If at A_{260}/A_{280} the purity of DNA is out of the 1.8–2.0 range, then the DNA should be purified to remove contaminants. Absorbance measurements at wave lengths other than 260 nm are used for determination of DNA purity. The relevant spectrum for this purpose lies between 320 and 220 nm. Any absorbance at 320 nm indicates contamination of particular nature. Proteins absorb maximally 280 nm due to the presence of tyrosine, phenylalanine, and tryptophan and absorption at this wavelength is used for detection of proteins in DNA samples. This is usually done by determination of the A_{260}/A_{280} ratio [8].

1.2 Materials

1.2.1. Bacterial (*E. coli*) DNA Isolation

1. Luria Bertani (LB) Broth
2. TE buffer—50 mM Tris, 50 mM EDTA (pH 8.0)
3. Tris (pH 8.0)—250 mM
4. Lysozyme—10 mg/ml
5. SDS—0.5 %
6. EDTA—0.4 M
7. Proteinase K—1 mg/ml
8. Phenol equilibrated with Tris (Phenol is a hazardous organic solvent. Always use suitable laboratory gloves when handling phenol containing solutions. Specific waste procedures may be required for the disposal of phenol containing solutions.)

9. Sodium acetate (pH 5.8)
10. Ethanol—95 %
11. RNase—200 µg/ml
12. Chloroform

**1.2.2. Gram – ve
Bacterial DNA Isolation
by CTAB Method**

1. Nutrient broth—25 ml
2. Tris EDTA (pH 8.0) (10 mM Tris-Cl, 1 mM EDTA)
3. 10 % SDS
4. 20 mg/ml proteinase K
5. 5 M NaCl
6. CTAB/NaCl Solution—Dissolve 4.1 g NaCl in 30 ml water and slowly add 10 g cetyltrimethylammonium bromide (CTAB) while stirring. If necessary, heat to 65 °C. Adjust to 100 ml
7. Chloroform
8. Isoamyl Alcohol
9. Buffered Phenol (8-Hydroxyquinoline, Liquefied phenol redistilled, 50 mM Tris-Cl, pH 8.0, TE buffer pH 8.0): Add 0.5 g of 8-hydroxyquinoline to a 2 l glass beaker. Gently add 500 ml liquefied phenol (crystals of redistilled phenol melted in a 65 °C bath). The phenol will turn yellow due to the 8-hydroxyquinoline, which is added as an antioxidant. Add 500 ml of 50 mM Tris base, cover with aluminum foil, and stir 10 min at low speed. Let phases separate at room temperature and gently decant the top (aqueous) phase into a suitable waste receptacle. Remove any residual aqueous phase with a glass pipette. Repeat twice with 500 ml each of 50 mM Tris-Cl, pH 8.0. Check pH of phenol with pH paper and repeat equilibration until pH = 8.0 and store at 4 °C in brown glass bottles or in clear glass bottles wrapped in aluminum foil
10. Isopropanol
11. 70 % (v/v Ethanol)

**1.2.3. Plasmid DNA
Isolation**

1. Luria Bertani (LB) medium supplemented with proper antibiotic
2. Lysis buffer I
 - (a) 25 mM Tris-Cl (pH 8.0)
 - (b) 50 mM Glucose
 - (c) 10 mM EDTA (pH 8.0)
 - (d) 0.2 mg/ml RNase A

3. Lysis buffer II (always freshly prepared)
 - (a) 0.2 NaOH
 - (b) 1% (w/v) SDS
4. Lysis buffer III
 - (a) 3 M potassium acetate (pH 5.5)
5. Chloroform: isoamyl alcohol (24:1)
6. Isopropanol
7. Ethanol (70 %)
8. TE buffer (pH 8.0)

1.2.4. *Actinomycetes* DNA Isolation

1. GYM broth
 - (a) 4.0 g Glucose, 4.0 g Yeast extract, 10.0 g Malt extract, 1 L Distilled water, pH 7.4
2. SET buffer
 - (a) 75 mM NaCl, 25 mM EDTA, 20 mM Tris, pH 7.5
3. SDS (10 %)
4. Lysozyme (10 mg/ml)
5. Proteinase K (20 mg/ml)
6. Rnase A (10 mg/ml)
7. 5 M NaCl
8. Phenol
9. Chloroform
10. Isoamyl alcohol
11. Isopropanol
12. Ethanol
13. Sodium acetate 3 M (pH 5.2)

1.2.5. *Yeast* DNA Isolation

1. Yeast extraction buffer A: 2 % Triton X-100, 1 % sodium dodecyl sulfate, 100 mM NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA pH 8.0, Phenol: chloroform: isoamylalcohol: phenol is presaturated with 10 mM Tris-HCl, pH 7.5.
2. Prepare a mixture of 25:24:1 phenol:chloroform:isoamyl alcohol (v/v/v). This solution can be stored at room temperature for up to 6 months, shielded from light.
3. Glass beads, diameter range 0.04–0.07 mm. Suspended as 500 mg/ml slurry in distiller water.
4. Ammonium acetate (4 M).

1.2.6. Fungal DNA Isolation by CTAB Method

1. CTAB extraction buffer: 0.1 M Tris-HCl, pH 7.5, 1 % CTAB (mixed hexadecyltrimethylammonium bromide), 0.7 M NaCl, 10 mM EDTA, 1 % 2-mercaptoethanol. Add proteinase K to a final concentration of 0.3 mg/ml prior to use.
2. Chloroform:isoamyl alcohol (24:1).

1.2.7. Purification of DNA by Phenol Extraction and Ethanol Precipitation

1. Phenol
2. TE buffer, pH 8.0 (10 mM Tris-HCl, pH 8.0; 1 mM EDTA, pH 8.0)
3. 24:1 (v/v) chloroform-isoamyl alcohol
4. 3 M potassium acetate, pH 5.5, prepared by adding glacial acetic acid to 3 M potassium acetate until this pH is obtained (store at 4 °C)
5. Cold 100 % ethanol (−20 °C)
6. Cold 70 % ethanol in sterile dH₂O (−20 °C)

1.2.8. Drop Dialysis Method

1. Drop dialysis filter
2. Sterile dialysis buffer (TE pH 8.0)
3. Petri dish

1.2.9. Purification on Sephacryl S-500 Spin Columns

1. Sephacryl S-500 column
2. 1×TM buffer
3. 100 mM Tris-HCl (pH 8.0)

1.2.10. DNA Fragment Purification from Agarose or Acrylamide

1. Crush and Soak Solution
 - (a) 500 mM NH₄OAc 3.3 g NH₄OAc
 - (b) 0.1 % SDS 0.1 g SDS
 - (c) 0.1 mM EDTA 20 ml 500 mM EDTA
Make up to 100 ml with Milli-Q and store at room temperature
2. 3 M NaOAc pH 5.2
 - (a) 24.6 g anhydrous sodium acetate pH to 5.2 with acetic acid and bring up to 100 ml with Milli-Q store at room temperature
3. Ethanol
4. Ethidium bromide (EtBr)
5. Phenol (do not expose to light)
6. Chloroform (store in brown bottle)
7. Other Reagents
 - (a) DMCS-treated glass wool (50 g)
 - (b) 0.22 mm disposable micro tip filters (syringe type) blue tips with melted tips to serve as pestle for crushing acrylamide

**1.2.11. DNA
Quantification and
Estimation by Gel
Electrophoresis**

1. Genomic DNA samples
2. Lambda HindIII DNA ladder
3. Loading dye

**1.2.12. DNA
Quantification and
Estimation by
Spectrophotometer**

1. Genomic DNA
2. Spectrophotometer
3. Cuvettes
4. Distilled water

1.3 Method

**1.3.1. Bacterial (*E. coli*)
DNA Isolation**

1. Grow the bacterial cells in 500 ml of LB broth medium
2. Subject the above overnight culture to centrifugation to obtain a pellet and dissolve the pellet in 5 ml of TE buffer [50 mM Tris (pH 8.0), 50 mM EDTA]
3. Freeze the above cell suspension at -20°C
4. To the frozen suspension, add 0.5 ml of 250 mM Tris (pH 8.0) and lysozyme (10 mg/ml) and thaw the contents at room temperature. After thawing again place on the ice for 45–50 min
5. Add 1 ml of 0.5 % SDS, 50 mM Tris (pH 7.5), 0.4 M EDTA, 1 mg/ml proteinase-K. Incubate in water bath at 50°C for 60 min
6. After incubation, extract with 6 ml of phenol and centrifuge at $10,000 \times g$ for 5 min
7. Transfer top layer to a new tube
8. Add 0.1 volume of 3 M Na-acetate and mix gently
9. Add 2 volumes of 95 % ethanol and mix by gentle inversion
10. Spool out the DNA so precipitated and add 5 ml of 50 mM Tris (pH 7.5), 1 ml of EDTA, 200 $\mu\text{g}/\text{ml}$ RNase. Dissolve it overnight by rocking at 4°C
11. Extract with equal volume of chloroform (mix by gentle inversion) and centrifuge at $10,000 \times g$ for 5 min
12. Transfer top layer to new tube
13. Add 0.1 volume of 3 M Na acetate and mix gently
14. Add 2 volumes of 95 % ethanol and gently mix the contents
15. Spool out the DNA and dissolve in 2 ml of TE buffer
16. Check the purity of the DNA and store at 4°C in TE till further use

1.3.2. Gram – ve Bacterial DNA Isolation by CTAB Method

1. Grow a 5 ml bacterial culture until saturated. Microcentrifuge 1.5 ml for 2 min or until a compact pellet forms. Resuspend pellet in 567 μ l TE buffer.
2. Add 30 μ l of 10 % SDS and 3 μ l of 20 mg/ml proteinase-K, mix thoroughly, and incubate 1 h at 37 °C.
3. Add 100 μ l of 5 M NaCl and mix thoroughly.
If NaCl concentration is <0.5 M, the nucleic acid may also precipitate.
4. Add 80 μ l of CTAB/NaCl solution, mix thoroughly, and incubate 10 min at 65 °C.
5. Add 1 volume (0.7–0.8 ml) of 24:1 chloroform/isoamyl alcohol, mix thoroughly, and microcentrifuge 4–5 min. Transfer supernatant to a fresh tube. If it is difficult to remove the supernatant, remove the interface first with a sterile toothpick.
6. Add 1 volume of 25:24:1 phenol/chloroform/isoamyl alcohol, extract thoroughly, and microcentrifuge 5 min transfer supernatant to a fresh tube.
7. Add 0.6 volume isopropanol and mix gently until a stringy white DNA precipitate forms. Transfer pellet to a fresh tube containing 70 % ethanol using a hooked, sealed Pasteur pipette. Alternatively, microcentrifuge briefly at room temperature, discard supernatant, and add 70 % ethanol to pellet.
8. Microcentrifuge 5 min at room temperature and dry pellet briefly in a lyophilizer. Resuspend in 100 μ l TE buffer.
Typical yield is 5–20 μ g DNA/ml starting culture (10^8 – 10^9 cells/ml).

1.3.3. Plasmid DNA Isolation

1. Prepare a small overnight culture of the host with plasmid in 2 ml of a rich medium (LB) containing the appropriate antibiotic at 37 °C with vigorous shaking.
2. Pour 1.5 ml of the above culture in a microfuge tube. Centrifuge in a microfuge for 30 s at maximum speed.
3. Discard the supernatant and dry the bacterial pellet. To this add 100 μ l of ice cold lysis buffer-I. Vortex vigorously.
4. Add 200 μ l of freshly prepared lysis solution-II to the above suspension. Mix the contents well by inverting the tubes several times.
5. Add 150 μ l of ice cold lysis solution III. Mix the contents by inverting the tubes several times.
6. Centrifuge the above lysate at maximum speed for 5 min.
7. Discard the pellet (chromosomal DNA) and collect the supernatant in fresh tube.

8. To the supernatant add an equal volume of chloroform:isoamyl alcohol. Mix and centrifuge at maximum speed for 2–3 min. Transfer the upper aqueous layer to a fresh tube.
9. Precipitate the plasmid DNA from the supernatant by adding equal volume of ice chilled isopropanol.
10. Centrifuge at maximum speed for 5 min.
11. Remove the supernatant and wash the pellet with 70 % of ethanol several times. Take care of the pellet as it may not adhere tightly to the tube.
12. Keep the microcentrifuge tube open so that ethanol is evaporated and pellet is dry.
13. Dissolve the pellet in 50 μ l of TE and store at -20°C .

1.3.4. *Actinomycetes* DNA Isolation

1. Grow the mycelia (1–2 ml) in a GYM broth shake culture at 37°C for 48 h in an orbital shaker at 120 rpm speed.
2. Centrifuge the broth at 800 rpm for 10 min and wash the pellet with sterile distilled water atleast twice followed by centrifugation.
3. To the mycelial pellet add 4 ml SET buffer. Add Lysozyme to a concentration of 1 mg/ml and incubate at 37°C for 1 h.
4. Add 0.1 volumes of 10 % SDS and 0.5 mg/ml proteinase-K and incubate at 55°C with occasional inversion for 2 h.
5. Add one-third volume 5 M NaCl and 1 volume of chloroform and incubate at room temperature for 0.5 h with frequent inversion.
6. Centrifuge the mixture at 8,000 rpm for 15 min and transfer the aqueous phase to a new tube using a blunt-ended pipette tip.
7. Precipitate the chromosomal DNA by the addition of 1 volume of isopropanol with gentle inversion.
8. Transfer the DNA to a new tube, rinse with 70 % ethanol, dry under vacuum, and dissolve in 100 μ l of sterile distilled water.
9. Treat the dissolved DNA with 20 mg/ml RNase-A at 37°C for 1 h.
10. Extract the samples with equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) and precipitate with 2.5 volume of ice cold ethanol and 0.1 volume of 3 M sodium acetate.
11. Wash the pellets with 70 % ethanol, vacuum dry, and dissolve in 100 μ l of sterile distilled water. Store the vials at -20°C .
12. Check the purity of DNA by agarose gel electrophoresis and quantify using spectrophotometer.

1.3.5. Yeast DNA Isolation

1. Collect cells from fresh 5 ml culture by centrifugation at $2,000 \times g$ for 10 min and resuspend in 0.5 ml of water.
2. Transfer cells to 1.5-ml microfuge tube and collect by centrifugation at $15,000 \times g$ for 10 min. Pour off supernatant and resuspend in residual liquid.
3. Add 0.2 ml of buffer A, 200 μ l of glass beads, and 0.2 ml of phenol:chloroform:isoamyl alcohol (25:24:1).
4. Vortex for 3 min and add 0.2 ml of TE.
5. Centrifuge at $15,000 \times g$ for 5 min and then transfer aqueous to new tube.
6. Add 1 ml of 100 % EtOH (room temperature), invert tube to mix, and centrifuge at $15,000 \times g$ for 2 min.
7. Discard supernatant and resuspend pellet in 0.4 ml of TE (no need to dry pellet).
8. Add 10 μ l of 4 M ammonium acetate, mix, and then add 1 ml of 100 % EtOH and mix.
9. Centrifuge at $15,000 \times g$ for 2 min and dry pellet. Resuspend in 50 μ l of TE.

1.3.6. Fungal DNA Isolation by CTAB Method

1. Grind 0.2–0.5 g (dry weight) of lyophilized mycelar pad in a mortar and pestle.
2. Transfer to a 50-ml disposable centrifuge tube.
3. Add 10 ml (for a 0.5 g pad) of CTAB extraction buffer.
4. Gently mix to wet the entire powdered pad.
5. Place in 65 °C water bath for 30 min.
6. Cool and add an equal volume of chloroform/isoamyl alcohol (24:1).
7. Mix and centrifuge at $2,000 \times g$ for 10 min at room temperature.
8. Transfer aqueous supernatant to a new tube.
9. Add an equal volume of isopropanol.
10. High molecular weight DNA should precipitate upon mixing and can be spooled out with a glass rod or hook.
11. Rinse the spooled DNA with 70 % ethanol.
12. Air dry, add 50 μ l of TE containing 20 μ g/ml RNase A. To resuspend the samples, place in 65 °C bath, allow pellets to resuspend overnight at 4 °C.

1.3.7. Phenol Extraction and Ethanol Precipitation

1. Add an equal volume of phenol to the DNA containing reaction mixture and vortex gently.
2. Separate the aqueous phase which contains the DNA from the organic phase by centrifugation in the microfuge, at 2,000 rpm for 5 min or at 8,000 rpm for 1 min.

3. Remove the aqueous phase with care into a fresh microfuge tube and add an equal amount of 24:1 (v/v) chloroform–isoamyl alcohol.
4. In order to precipitate the DNA, add a 0.1 volume of 3 M sodium acetate, pH 5.5, to the aqueous phase and then 2 volumes of absolute ethanol. Incubate at -20°C overnight or for shorter periods at -80°C (e.g., 20–30 min).
5. Recover the precipitated DNA by centrifugation in the microfuge at 10,000 rpm for 5–15 min. Remove the ethanol with care and dry the pellet in a desiccator or 50°C oven for 5 min. An extra wash with 70 % (v/v) ethanol may be included to remove excess salt from the pellet. The dried DNA may be resuspended in sterile TE, pH 8.0, or water and stored at 4°C for further manipulation or at -20°C for long-term storage.
6. This procedure denatures and removes contaminating protein from a DNA sample.

1.3.8. Drop Dialysis Method

1. Gently place a drop dialysis filter, floating correct-side up, on 10–20 ml of sterile dialysis buffer (TE, pH 8.0, or water) in a Petri dish.
2. Gently pipette the DNA sample (10–100 μl) onto the filter.
3. Allow to dialyze for 1–2 h before removing the DNA for further analysis.

1.3.9. Purification on Sephacryl S-500 spin columns

1. Thoroughly mix a fresh new bottle of Sephacryl S-500, distribute in 10 ml portions, and store in screw cap bottles or centrifuge tubes in the cold room.
2. Prior to use, briefly vortex the matrix and without allowing to settle, add 500 μl of this slurry to a mini-spin column (Millipore) which has been inserted into a 1.5-ml microcentrifuge tube.
3. Following centrifugation at 2,000 rpm in a table top centrifuge, carefully add 200 μl of 100 mM Tris–HCl (pH 8.0) to the top of the Sephacryl matrix and centrifuge for 2 min. at 2,000 rpm. Repeat this step twice more. Place the Sephacryl matrix-containing spin column in a new microcentrifuge tube.
4. Then, carefully add 40 μl of DNA to the Sephacryl matrix (saving 2 μl for later agarose gel analysis) and centrifuge at 2,000 rpm for 5 min. Remove the column, save the solution containing the eluted, large DNA fragments (fraction 1). Apply 40 μl of $1\times\text{TM}$ buffer and recentrifuge for 2 min at 2,000 rpm to obtain fraction 2 and repeat this $1\times\text{TM}$ rinse step twice more to obtain fractions 3 and 4.
5. To check the DNA purity, load 3–5 μl of each eluant fraction onto a 0.7 % agarose gel that includes as controls, 2 μl of DNA saved from step 4 above.

1.3.10. DNA Fragment Purification from Agarose or Acrylamide

1.3.10.1. Agarose Gels

1. Prepare spin columns by cutting off the cap of a 0.5 ml eppendorf tube and forming a hole in the bottom with a hot 18 gauge needle. Fill this “mini-column” with a small ball of DMCS-treated glass wool and pack down with a pipet tip.
2. Cut out the desired band from an agarose gel and place in a spin column inside a 1.5-ml eppendorf tube with the top cut off.
3. Spin at 6,000 rpm in a microfuge for 10 min.
4. Phenol/chloroform extract flow through and EtOH precipitate with glycogen or tRNA and 10 % v/v of 3 M NaOAc, pH 5.2.
5. Wash and dry, resuspend in 20 μ l TE, run 10 μ l on a gel, and use 1–2 μ l for a ligation.

1.3.10.2. Acrylamide Gels

1. Run a 4–6 % acrylamide gel in 1 \times TBE, stain in EthBr (1–10 mg/ml), and cut out the desired band.
2. Crush the acrylamide with a p1000 tip with a melted end to resemble a pestle for the eppendorf “mortar.”
3. Add 1 ml crush and soak solution and incubate overnight at 37 °C.
4. Spin in the microfuge for 10 min at 14,000 rpm. Remove as much liquid as possible and add another 500 μ l of crush and soak solution.
5. Repeat the spin and pool the recovered supernatant.
6. Add 0.1 volumes of 3 M NaOAc, 2.5 volumes of EtOH, and carrier (see above).
7. Spin as usual, wash, and dry. Resuspend in 20 μ l TE.

1.3.11. DNA Quantification and Estimation by Gel Electrophoresis

1. Obtain ice bucket and keep DNA samples on ice
2. Heat *Hind* III ladder at 60–65 °C for 3 min
3. Place on ice
4. Cut small parafilm piece
5. Pipet 3 μ l dots loading dye onto parafilm
6. Pipet 1 μ l each DNA sample onto loading dye
7. Pipet 1 μ l HindIII DNA onto loading dye dot
8. Record DNA positions in lab notebook
9. Load gel
10. Store remaining DNA at –20 °C

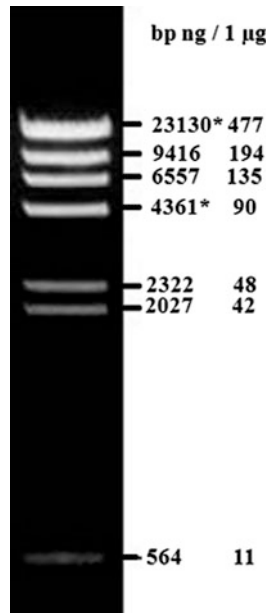


Fig. 1.2 Determine the DNA concentration by comparing with the band intensity

11. Plug gel box into power supply
12. Set voltage to ~80 V
13. Run gel until loading dye reaches approximately 3/4 th of the gel length
14. Visualize and photograph in gel doc room
15. Dispose of gel
16. By comparing the brightness of the bands, you can estimate your DNA concentration. For instance, if your band is in halfway between the 4,361 bp and 2,322 bp fragments, you could estimate that your concentration is ~70 ng/1 µg
17. Alternately, you can determine the concentration by calculations. For example, if your band's brightness seems similar to the brightness of the 4,361 band (Fig. 1.2), you perform the following calculation:

$$\frac{1 \mu\text{g}/L}{48,502 \text{ bp}} \times 4,361 \text{ bp} \times \frac{1 \text{ ng}}{1,000 \mu\text{g}} = X \text{ ng}/\mu\text{l}$$

1.3.12. DNA Quantification and Estimation by Spectrophotometer

1. Use H₂O or 1 × TE as a solvent to suspend the nucleic acids and place each sample in a quartz cuvette.
2. Zero the spectrophotometer with a sample of solvent.
3. For more accurate readings of the nucleic acid sample of interest, dilute the sample to give readings between 0.1 and 1.0.

4. For a 1-cm path length, the optical density at 260 nm (OD_{260}) equals 1.0 for the following solutions:
 - (a) 50 $\mu\text{g}/\text{ml}$ solution of dsDNA
 - (b) 33 $\mu\text{g}/\text{ml}$ solution of ssDNA
 - (c) 20–30 $\mu\text{g}/\text{ml}$ solution of oligonucleotide
 - (d) 40 $\mu\text{g}/\text{ml}$ solution of RNA
5. Contamination of nucleic acid solutions makes spectrophotometric quantitation inaccurate.
6. Calculate the OD_{260}/OD_{280} ratio for an indication of nucleic acid purity.
7. Pure DNA has an OD_{260}/OD_{280} ratio of ~ 1.8 ; pure RNA has an OD_{260}/OD_{280} ratio of ~ 2.0 .
8. Low ratios could be caused by protein or phenol contamination.

1.3.13. Calculation

A sample of dsDNA was diluted 50 \times . The diluted sample gave a reading of 0.65 on a spectrophotometer at OD_{260} . To determine the concentration of DNA in the original sample, perform the following calculation:

1. dsDNA concentration = 50 $\mu\text{g}/\text{ml}$ \times OD_{260} \times dilution factor
2. dsDNA concentration = 50 $\mu\text{g}/\text{ml}$ \times 0.65 \times 50
3. dsDNA concentration = 1.63 mg/ml

References

1. Sambrook J, Russell D (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, New York
2. Bruce B, Eric DG, Sue K, Richard MM, Jane R (1999) Genome analysis: a laboratory manual, vol I. Cold Spring Harbor Laboratory Press, New York
3. Harju S, Fedosyuk H, Peterson KR (2004) Rapid isolation of yeast genomic DNA: Bust n' Grab. BMC Biotechnol 21:4–8
4. Lee SB, Taylor JW (1990) Isolation of DNA from fungal mycelia and single spores. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a guide to methods and applications. Academic, San Diego, CA, pp 282–287
5. Frederick MA, Roger B, Robert EK, David DM, Seidman JG, John AS, Kelvin S (2002) Short protocols in molecular biology (Vol-I): a laboratory manual (5th edn). Wiley, USA
6. Carter MJ, Milton ID (1993) An inexpensive and simple method for DNA purifications on silica particles. Nucleic Acids Res 21:1044
7. Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheim Van Dillen PME, Van der Noordaa J (1990) Rapid and simple method for purification of nucleic acids. J Clin Microbiol 28: 495–496
8. Tataurov AV, You Y, Owczarzy R (2008) Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids. Biophys Chem 133:66–70

Chapter 2

Fluorescent-Based Detection, Quantitation, and Expression of Viral Gene by qRT-PCR

Shelly Praveen and Vikas Koundal

Abstract

Using fluorescent reporter molecules, viral gene(s) can be quantified for diagnostics as well as for gene expression studies by quantitative PCR. The process is based on the detection of the fluorescence produced by a reporter molecule which increases, as the reaction proceeds. This occurs due to the accumulation of the PCR product with each cycle of amplification. The procedure follows the general principle of polymerase chain reaction; its key feature is that the amplified DNA is quantified as it accumulates in the reaction in real time after each amplification cycle. Here we discuss the detailed explanation of various fluorescent molecules and strategies to determine the viral load. These experiments are equally efficient in determining viral gene expression studies.

2.1 Introduction

Over last several years, the development of novel chemistries and instrumentation platforms enabling detection of PCR products on real-time basis has lead to wide spread adoption of quantitative PCR (Q-PCR/qPCR) as the method of choice for diagnostic and analyzing changes in gene expression [1]. It is called “real-time PCR,” because it allows us to monitor the increase in the amount of DNA as it is amplified. Real-time polymerase chain reaction is also known as kinetic polymerase chain reaction, because it is the most sensitive technique to amplify and simultaneously quantify a targeted DNA molecule compared to the commonly used Northern and Southern blotting techniques. It enables both detection and quantification (as absolute number of copies or relative amount when normalized to DNA input or additional normalizing genes) of a specific sequence in a DNA sample. Q-PCR can be used to quantify RNA levels from much smaller samples. In fact, this technique is sensitive enough to enable quantitation of RNA from a single cell.

Real-time PCR or Q-PCR is a variation of the standard PCR technique used to quantify DNA or messenger RNA (mRNA) in a sample. Quantification of amplified product is obtained using fluorescent probes and specialized machines that measure fluorescence while performing temperature changes needed for the PCR cycles. Real-time PCR is based on the detection of the fluorescence produced by a reporter molecule which increases, as the reaction proceeds. This occurs due to the accumulation of the PCR product with each cycle of amplification. These fluorescent reporter molecules include dyes that bind to the double-stranded DNA (i.e., SYBR Green) or sequence-specific probes. The procedure follows the general principle of polymerase chain reaction; its key feature is that the amplified DNA is quantified as it accumulates in the reaction in real time after each amplification cycle. Using sequence-specific primers, the relative number of copies of a particular DNA or RNA sequence can be determined. We use the term relative since this technique tends to be used to compare relative copy numbers between tissues, organisms, or different genes relative to a specific housekeeping gene (or reference gene). Housekeeping genes are constitutively expressed at a relatively constant level in cells, and they are always present in all known conditions. To obtain the value of the target gene under investigation and the value of the housekeeping gene in the same sample, a standard curve can be used. In this, the absolute concentration of the target gene is divided by the absolute concentration of the housekeeping gene. The resulting target/reference ratio that expresses the amount of target gene is then normalized to the level of the reference gene within each unknown sample [2]. The quantification arises by measuring the amount of amplified product at each stage during the PCR cycle. DNA/RNA from genes with higher copy numbers will appear after fewer melting, annealing, extension PCR cycles.

We present here how to quantitate 2b gene (RNAi suppressor) of *Cucumber mosaic virus* (CMV) in different plant samples, using fluorescent molecules [3]. CMV is the type member of the *Cucumovirus* genus, in the family *Bromoviridae*. Cucumber mosaic, first described in 1916, was one of the earliest plant diseases attributed to a virus. CMV genome consists of positive sense, single-stranded RNA. CMV encodes five proteins, distributed on three genomic RNAs, i.e., RNA1 which is the only monocistronic RNA, encoding the 1a protein that is required for viral replication and contains methyl-transferase and helicase motifs. RNA2 encodes the 2a protein, the viral polymerase, and the 2b protein, the RNAi suppressor. RNA3 encodes the movement protein (MP), and the coat protein (CP) expressed from the subgenomic RNA4 and the satellite RNAs (satRNAs). The subgenomic RNA4 and the satellite RNAs of CMV are small linear RNA that does not carry any apparent coding capacity (Fig. 2.1).

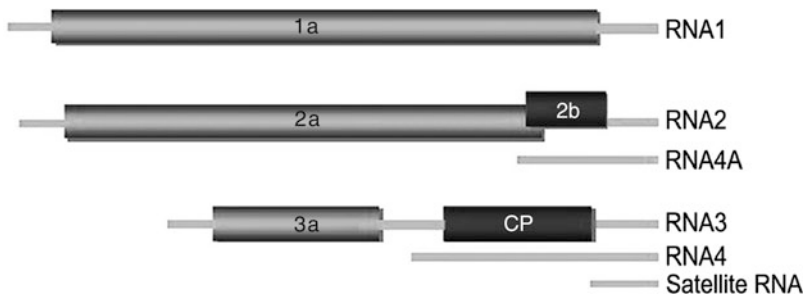


Fig. 2.1 CMV genome consists of single-stranded positive sense RNA.

CMV infects over 1,000 species of hosts, including members of 85 plant families, making it the broadest host range virus known. Tomatoes infected with the *Cucumber mosaic virus* develop a slight yellowing and mottling of the older leaves. The expanding leaves typically become twisted, curl downward, and develop a “shoestring” appearance as a result of a restriction of the leaf surface to a narrow band around the midrib of the leaf. Diseased plants are stunted and produce poor fruit yield.

2.2 Materials

2.2.1. Instrument and Set Up

- 1. Real-time PCR thermal cycler; Light Cycler[®] 480 II (Roche).
- 2. Clear LightCycler[®] 480 multiwell plates.
- 3. LightCycler[®] 480 Sealing Foil.
- 4. 2× Light CyclerR 480 SYBR Green I master mix/TaqMan probe.
- 5. Gene-specific primers.

2.2.2. QRT-PCR Reaction Requirements (Using TaqMan Probe)

Total RNA (as CMV is a RNA virus) from the infected plant leaf samples

1. RNA (10–100 ng)	30.0 µl
2. 10× TaqMan buffer	5.0 µl
3. MgCl ₂ (25 mM)	5.0 µl
4. dNTPs (10 mM)	2.0 µl
5. Primer F (10 µM)	2.5 µl
6. Primer R (10 µM)	2.5 µl
7. TaqMan probe (10 µM)	1.0 µl
8. Taq Polymerase (5 U)	0.5 µl
9. M-MuLV Reverse Transcriptase (20 U)	0.5 µl
10. RNase inhibitor (20 U)	1.0 µl

2.2.3. Cycling

Parameters

1. Reverse transcription (using M-MuLV) at 48 °C for 30 min.
2. *Taq* activation 95 °C for 10 min.

PCR profile:

1. Denaturation at 95 °C for 15 s.
2. Annealing/extension at 60 °C for 1 min (repeated 40 times).

2.2.4. Components

2.2.4.1. Primers and Probe

Whenever possible, primers and probes should be selected in a region with a G/C content of 30–80 %. Regions with G/C content in excess may not denature well during thermal cycling, leading to a less efficient reaction. In addition, G/C-rich sequences are susceptible to nonspecific interactions that may reduce reaction efficiency and produce nonspecific signal in SYBR Green assays. For this same reason, primer and probe sequences containing repeats of four or more G bases should be avoided. A/T-rich sequences require longer primer and probe sequences in order to obtain the optimum melting temperatures. This is rarely a problem for quantitative assays; however, probes approaching 40 base pairs can exhibit less efficient quenching and produce lower synthesis yields. Primer should be highly purified ideally; HPLC purified primers should be used and their concentration should be in the range of 0.3–1 μ M, ideally 0.5 μ M. The last five bases on the 3' end of the primers should contain no more than two C and/or G bases, which is another factor that reduces the possibility of nonspecific product formation. Under certain circumstances, however, such as a G/C-rich template sequence, this recommendation may have to be relaxed to keep the amplicon under 150 base pairs in length. It should be followed as often as possible, and even when it is not possible, primer 3' ends extremely rich in G and/or C bases should be avoided. The T_m of primers is adjusted in the range of 58–60 °C as both the annealing and extension step are achieved in a single step of real-time PCR.

2.2.4.2. Probe Selection Criteria

1. Select the probe first and design the primers as close as possible to the probe without overlapping it.
2. Keep the G/C content in the 30–80 % range.
3. Avoid runs of an identical nucleotide, especially for guanine, where repeats of four or more should be avoided, and there should be no G on the 5' end.
4. T_m of the probe should be in the range of 60–70 °C.
5. Select the probe with more C compared to G bases.

Selecting primers and probes with the recommended T_m is one of the factors that allow the use of universal thermal cycling parameters. Having the probe T_m 8–10 °C higher than that of the

primers ensures that the probe is fully hybridized during primer extension. The required parameters for well-designed primers and probe have been well documented. These parameters include T_m for the probe that is 10 °C higher than the primers, primer T_m between 58 and 60 °C, amplicon size between 50 and 150 bases, and absence of 5' Gs.

2.2.4.3. Template

A critical aspect of performing real-time PCR is to begin with a template that is of high purity. The DNA should be about 5–30 ng in concentration ideally; 25 ng DNA template is used in 20 µl reaction mix. Size of amplicon should be <500 bp. Small amplicons are favored because they promote high-efficiency assays. In addition, high-efficiency assays enable relative quantification to be performed using the comparative method or threshold cycle (C_t). This method increases sample throughput by eliminating the need for standard curves when looking at expression levels of a target relative to a reference control.

2.2.5. General Recommendations for Real-Time RT-PCR

The optimal concentrations of the real-time PCR reagents are as follows:

1. Magnesium chloride concentration should be between 4 and 7 mM.
2. Concentrations of dNTPs in TaqMan reaction should be 200 µM of each dNTPs.

Typically 1.25 U of *Taq* DNA polymerase is used in a 50-µl reaction mixture. This is the minimum requirement; if necessary, optimization can be done by increasing this amount by 0.25 U increments.

2.3 Methods

2.3.1. Standard Curve Method

In this method, a standard curve is first plotted from DNA/RNA sample of known concentration. This curve is then used as a reference standard for extrapolating quantitative information for samples of unknown DNA/RNA concentration. Nucleic acids like DNA, RNA, in vivo generated ssDNA or any cDNA sample can be used to construct standard curve. For the standard curve first the standard sample is quantified accurately, spectrophotometrically and is then converted to copy number based on molecular weight of the sample used. In this method, a standard curve is first plotted from DNA/RNA sample of known concentration. This curve is then used as a reference standard for extrapolating quantitative information for samples of unknown DNA/RNA concentration. Nucleic acids like DNA, RNA, in vivo generated ssDNA or any cDNA sample can be used to construct standard

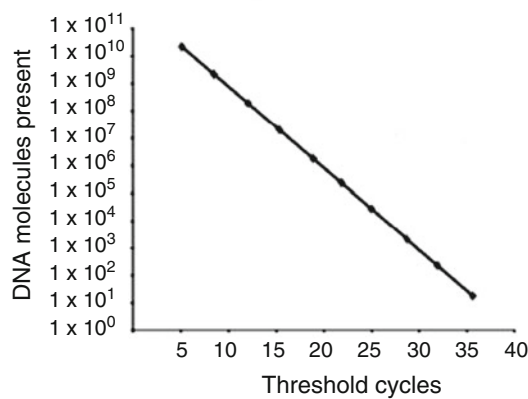


Fig. 2.2 Standard curve for absolute quantitation.

curve. For the standard curve first the standard sample is quantified accurately, spectrophotometrically and is then converted to copy number based on molecular weight of the sample used (Fig. 2.2).

Though RNA standards can be used, their stability can be a source of variability in the final analyses also; using RNA standards involves the construction of cDNA plasmids that have to be in vivo transcribed into the RNA standards. To check the variation introduced due to the variable RNA inputs, normalization can be done using a housekeeping gene.

2.3.2. Comparative Threshold (C_t) Method

This involves comparing the C_t values of sample of interest with a control or calibrator such as a non-treated sample or RNA from normal tissue. The C_t value of both the sample of interest and calibrator are normalized to an appropriate endogenous housekeeping gene.

The comparative C_t method is also known as $2^{-[\Delta]\Delta C_t}$ method, where:

$$[\Delta]\Delta C_t = [\Delta] C_{t, \text{sample}} - [\Delta] C_{t, \text{reference}}$$

Here, $[\Delta] C_{t, \text{sample}}$ is the C_t value for any sample normalized to the endogenous housekeeping gene and $[\Delta] C_{t, \text{reference}}$ is C_t value for the calibrator also normalized to the endogenous housekeeping gene.

For the $[\Delta]\Delta C_t$ calculation to be valid, the amplification efficiencies of the target and the endogenous reference must be approximately equal. This can be estimated by looking at how $[\Delta] C_t$ varies with template dilution. If the plot of cDNA dilution versus ΔC_t is close to zero, it implies that the efficiencies of the target and housekeeping genes are very similar. If a housekeeping gene cannot be found whose amplification efficiency is similar to the target, then the standard curve method is preferred (Fig. 2.3).

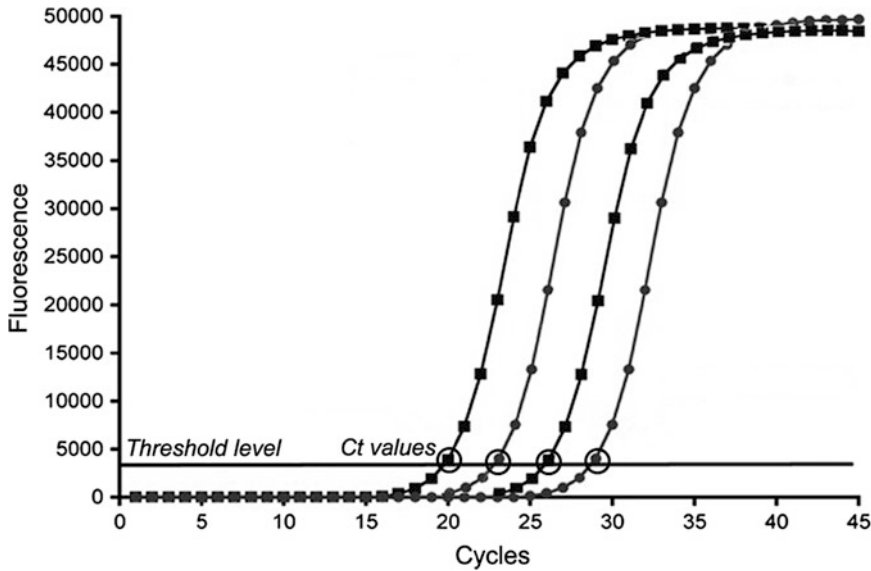


Fig. 2.3 Graph representing a typical curve with critical threshold value.

2.3.3. Reaction Set and Thermal Cycling

1. Prepare the reaction mix (according to the components given in Sect. 2.2.2) and load it on the multi-well plate.
2. Set up the cycling parameter (shown in Sect. 2.2.3 for CMV-2b gene).
3. Run the program and analyze the results.

2.4 Notes

The quantification is done by measuring the amount of amplified product at each stage during the PCR cycle. DNA/RNA from genes with higher copy numbers will appear after fewer PCR cycles. Quantification of amplified product is obtained using fluorescent probes and specialized machines that measure fluorescence while performing temperature changes required for the PCR cycles. Real-time PCR is based on the detection of the fluorescence produced by a reporter molecule which increases, as the reaction proceeds. This occurs due to the accumulation of the PCR product with each cycle of amplification. These fluorescent reporter molecules include dyes that bind to the double-stranded DNA (i.e., SYBR Green) or sequence-specific probes (TaqMan probes). The procedure follows the general principle of polymerase chain reaction; its key feature is that the amplified DNA is quantified as it accumulates in the reaction in real time after each amplification cycle.

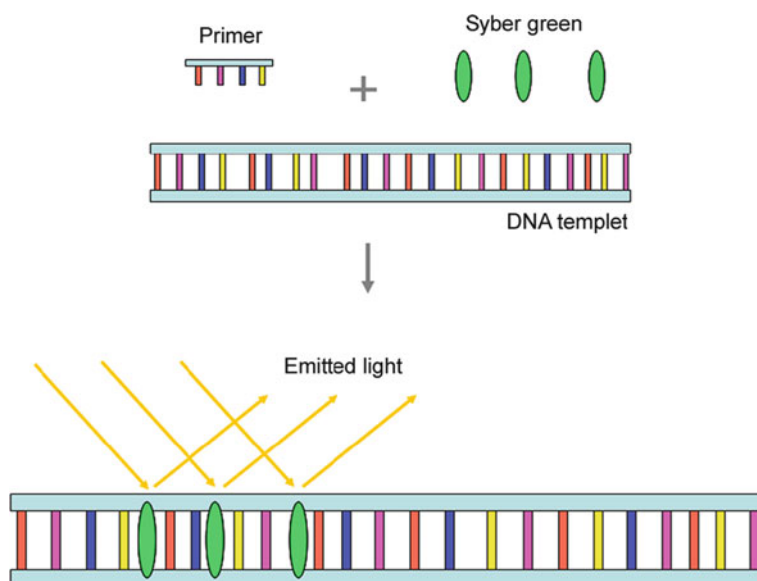


Fig. 2.4 SYBER Green; A fluorescent dye binds with double-stranded DNA.

2.4.1. SYBER Green

SYBER Green provides the simplest and most economical format for detection and quantitation of PCR products in real-time reactions. SYBER Green binds double stranded DNA and upon excitation emits light. An increase in DNA product during PCR leads to an increase in fluorescence intensity and is measured at each cycle, thus allowing DNA concentrations to be quantified (Fig. 2.4).

The advantages of SYBER Green are that it is inexpensive, easy to use, and sensitive. However, dsDNA dyes such as SYBR Green binds to all dsDNA PCR products, including nonspecific PCR products (such as “primer dimers”). This can potentially interfere with/or prevent accurate quantification of the intended target sequence. For single product PCR reactions with well-designed primers, SYBER Green can work extremely well, with spurious nonspecific background showing up in very late cycles.

1. The reaction is prepared as usual, with the addition of fluorescent dsDNA dye (instead of TaqMan probe as described in protocol).
2. The reaction is run in a thermocycler, and after each cycle, the levels of fluorescence are measured with a detector; the dye only fluoresces when bound to the dsDNA (i.e., the PCR product). With reference to a standard dilution, the dsDNA concentration in the PCR can be determined.

Like other real-time PCR methods, the values obtained do not have absolute units associated with it (i.e., DNA/RNA copies/cell). A comparison of a measured DNA/RNA sample to a



Fig. 2.5 The TaqMan probe. The *red circle* represents the quenching dye that disrupts the observable signal from the reporter dye (*green circle*) when it is within a short distance.

standard dilution will only give a fraction or ratio of the sample relative to the standard, allowing only relative comparisons between different tissues or experimental conditions. To ensure accuracy in the quantification, it is usually necessary to normalize expression of a target gene to stably expressed housekeeping genes, e.g., Ubiquitin. SYBR Green is the most widely used double-strand DNA-specific dye reported for real-time PCR. SYBR Green binds to the minor groove of the DNA double helix. In the solution, the unbound dye exhibits very little fluorescence. This fluorescence is substantially enhanced when the dye is bound to double-strand DNA. SYBR Green remains stable under PCR conditions and the optical filter of the thermocycler can be affixed to harmonize the excitation and emission wavelengths. Ethidium bromide can also be used in real-time PCR for detection of amplification but its carcinogenic nature renders its use restrictive.

2.4.2. Hydrolysis Probe

The hydrolysis probe chemistry relies on the 5′–3′ exonuclease activity of *Taq* polymerase, which degrades a hybridized non-extendible DNA probe [4] during the extension step of the PCR, e.g., TaqMan probes. TaqMan probe is designed to hybridize in a region within the amplicon and is dual labeled with a reporter dye and a quenching dye. The reporter dye is attached to the 5′ end of the probe and the quencher at the 3′ end. The close proximity of the reporter to quencher prevents detection of its fluorescence (Fig. 2.5).

During the annealing stage of the PCR, both probe and primers anneal to the DNA target. Polymerization of a new DNA strand is initiated by TaqMan polymerase from the primers, and once the polymerase reaches the probe, its 5′–3′ exonuclease activity degrades the probe, physically separating the fluorescent reporter from the quencher, resulting in fluorescence, and this fluorescence is detected and measured in the real-time PCR thermocycler. The more times the denaturing and annealing takes place, the more opportunities there are for the TaqMan probe to bind and, in turn, the more emitted light is detected, and its geometric increase corresponding to exponential increase of the product is used to determine the threshold cycle (C_t) in each reaction.

Fluorescent reporter probes are more accurate and reliable of the method. It uses a sequence-specific RNA or DNA-based probe to quantify only the DNA containing the probe sequence;

therefore, use of the reporter probe significantly increases specificity, and allows quantification even in the presence of some nonspecific DNA amplification. This potentially allows for multiplexing assaying for several genes in the same reaction by using specific probes with different colored labels, provided that all genes are amplified with similar efficiency. Well-designed TaqMan probes require very little optimization.

2.4.3. Hybridization Probe

The hybridization probes are used for DNA detection and quantitation, providing maximum specificity for product identification. Two specifically designed, sequence-specific oligonucleotide probes, labeled with different dyes, are used. The sequences of the probes are selected so that they can hybridize to the target sequences on the amplified DNA fragment in a head-to-tail orientation, thus bringing the two dyes into close proximity. The donor dye (fluorescein) is excited by the blue light source and emits green fluorescent light at a slightly longer wavelength. At close proximity, the energy emitted excites the acceptor dye attached to the second hybridization probe, which then emits fluorescent light at a different wavelength. The amount of fluorescence emitted is directly proportional to the amount of target DNA generated during the PCR reaction.

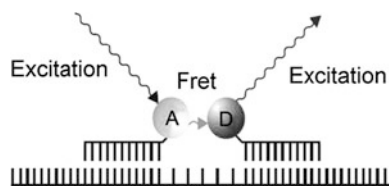
2.4.3.1. FRET Probes

Fluorescence resonance energy transfer (FRET) is transfer of energy from excited state, i.e., from the initially excited donor (D) to an acceptor (A). The hybridization probe system consists of two oligonucleotides labeled with different marker fluorescent dyes and these two probes designed to hybridize in close proximity on the target DNA. Interaction of the two dyes can only occur when both are bound to their target. The donor probe is labeled with fluorophore at the 3' end and the acceptor probe at 5' end. During PCR, the two different oligonucleotides hybridize to adjacent regions of the target DNA such that the fluorophores, which are coupled to the oligonucleotide, are in close proximity in the hybrid structure. The donor fluorophore is excited by an external light source and then passes part of its excitation energy to the adjacent acceptor fluorophore. The excited acceptor fluorophore emits light at a longer wavelength, which can then be detected and measured (Fig. 2.6). The light source cannot excite the acceptor dye.

Applications of FRET probes are in:

1. Quantitative PCR.
2. DNA copy number measurements.
3. Pathogen detection assays.
4. Single nucleotide polymorphism (SNP) genotyping.
5. Verification of microarray results.

Fig. 2.6 Fluorescence resonance energy transfer (FRET).



2.4.3.2. Molecular Beacons

Molecular beacons are hairpin-shaped oligonucleotide in their unhybridized state containing a fluorophore on one end and a quenching dye on the opposite end. Under conditions when probe is not hybridized to its complementary target, the fluorescent and quenching dye remain proximal to one another, thus preventing fluorescence resonance energy transfer (FRET). Whereas, when the probe encounters a target molecule, it forms a probe–target hybrid that is longer and more stable than the stem hybrid; this causes the fluorophore and the quencher to move away from each other and causes emission of fluorescence. Molecular beacons are designed so that their probe sequence is just long enough for a perfectly complementary probe–target hybrid to be more stable than the stem hybrid. The length of the probe sequence (10–40 nt) is chosen in such a way that the probe target hybrid is stable in the conditions of the assay. The stem sequence (5–7 nt) is chosen to ensure that the two arms hybridize to each other but not to the loop sequence (Fig. 2.7).

The computer program is used to predict melting temperature of the stem and also to predict whether the intended stem-and-loop conformation will occur or not. Molecular beacons can be synthesized that possess differently colored fluorophores, enabling assays to be carried out that simultaneously detect different targets in the same reaction.

Molecular beacons are thus ideal probes for use in diagnostic assays designed for genetic screening, SNP detection, and pharmacogenetic applications. In summary, molecular beacons have three key properties that enable the design of new and powerful diagnostic assays:

1. They only fluoresce when bound to their targets.
2. They can be labeled with a fluorophore of any desired color.
3. They are so specific that they easily discriminate single-nucleotide polymorphisms.

2.4.3.3. Scorpions

In Scorpion probes, sequence-specific priming and PCR product detection are achieved using a single oligonucleotide. The Scorpion probe maintains a stem–loop configuration in the unhybridized state. The fluorophore is attached to the 5' end and is quenched by a moiety coupled to the 3' end. The 3' portion of the stem also contains sequence that is complementary to the extension product of the primer. This sequence is linked to the



Fig. 2.7 Molecular beacon; a hairpin fluorescent probe.

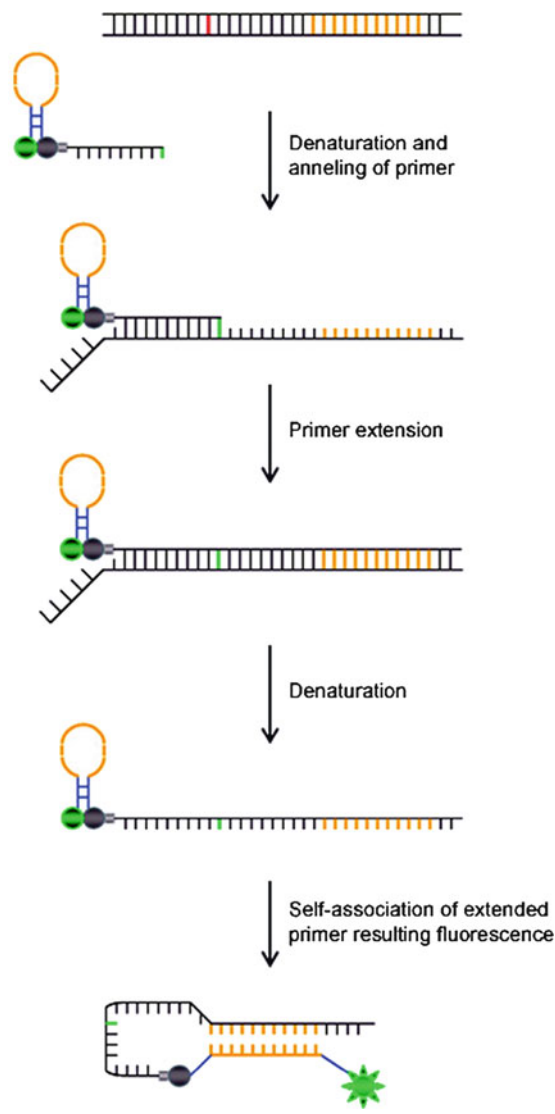


Fig. 2.8 Scorpion probe; a single oligonucleotide used in priming as well as in probing.

5' end of a specific primer via a non-amplifiable monomer. After extension of the Scorpion primer, the specific probe sequence is able to bind to its complement within the extended amplicon thus opening up the hairpin loop. This prevents the fluorescence from being quenched and a signal is observed (Fig. 2.8).

It is possible to choose between the open and closed Scorpion format. Closed format means that the probe part of the Scorpion is designed to have two stems at each end that are complimentary to each other so that it will be in a beacon-like (link) secondary structure when it is not yet hybridized to the primer's extension product. This way a fluor and quencher that are attached to the 5' and 3' ends of the probe are in close proximity to each other. Hence, when the Scorpion is free in solution no fluorescence can be detectable. When the Scorpion unfolds as the probe binds to the extended primer, the fluor and quencher will be separated and fluorescence can be detected as to quantify the amount of PCR product. In the open format, the probe part of the Scorpion does not have a specific secondary structure in the unhybridized form and contains a fluor. A separate quencher oligonucleotide is designed simultaneously. This quencher will bind to the probe part of the Scorpion when the Scorpion is not bound to its intended target so as to prevent fluorescence. As the Scorpion binds to the target, the quencher and probe will be separated from each other when the probe hybridizes to the extension product of the primer and hence, fluorescence can be detected and used to quantify the amount of PCR product.

References

1. VanGuilder HD, Vrana KE, Freeman WM (2008) Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* 44: 619–626
2. Dhanasekaran S, Mark Doherty T, Kenneth J (2010) Comparison of different standards for real-time PCR-based absolute quantification. *J Immunol Methods* 354:34–39
3. Praveen S, Mangrauthia SK, Singh P, Mishra AK (2008) Behavior of RNAi suppressors protein 2b of *Cucumber mosaic virus in planta* in presence of and absence of virus. *Virus Genes* 37: 96–102
4. Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl* 4:357–362

Chapter 3

Restriction Enzymes and Their Role in Microbiology

Sukumar Mesapogu, Chandra Mouleswararao Jillepalli,
and Dilip K. Arora

Abstract

The ability to cleave DNA at specific sites is one of the cornerstones of today's methods of DNA manipulation. Restriction endonucleases are intended to cleave duplex DNA at specific target sequences with the production of defined fragments. Restriction fragment length polymorphisms (RFLPs) are differences in genomic DNA sequences between individuals that are revealed by cleaving each individual's DNA with restriction enzymes, separating the DNA fragments according to size. Each enzyme cuts the palindrome at a particular site, and two different enzymes may have the same recognition sequence, but cleave the DNA at different points within that sequence. The cleavage sites fall into three different categories, either flush (or blunt) in which the recognition site is cut in the middle, or with either 5'- or 3'-overhangs, in which case unpaired bases will be produced on both ends of the fragment.

3.1 Introduction

Restriction endonucleases are a class of enzymes that cut DNA molecules. Each enzyme recognizes a unique sequence of nucleotides in the DNA strand, usually about 4–6 base pairs long. The sequences are palindromic in that the complimentary DNA strand has the same sequence only in the reverse direction, so both strands of DNA are cut at the same location [1].

Restriction enzymes are found in many different strains of bacteria where their biological role is to participate in cell defense. These enzymes “restrict” foreign (e.g., viral) DNA that enters the cell, by destroying it. Among the first of these “restriction enzymes” to be purified were *EcoRI* and *EcoRII* from *Escherichia coli*, and *HindII* and *HindIII* from *Haemophilus influenzae*. These enzymes were found to cleave DNA at specific sites, generating discrete, gene-size fragments that could be rejoined. Researchers were quick to recognize that restriction enzymes provided them with a remarkable new tool for investigating gene organization, function, and expression. The host cell has

a restriction-modification system that methylates its own DNA at sites specific for its respective restriction enzymes, thereby protecting it from cleavage. Over 800 known enzymes have been discovered that recognize over 100 different nucleotide sequences.

There are three different types of restriction enzymes. Type I cuts DNA at random locations as far as 1,000 or more base pairs from the recognition site. Type III cuts at ~25 base pairs from the site. Types I and III require ATP and may be large enzymes with multiple subunits. Type II enzymes, which are predominantly used in biotechnology, cut DNA within the recognized sequence without the need for ATP and are smaller and simpler. Type II restriction enzymes are named according to the bacterial species from which they are isolated [2]. For example, the enzyme EcoRI was isolated from *E. coli*.

Type II restriction enzymes can generate two different types of cuts depending on whether they cut both strands at the center of the recognition sequence or each strand closer to one end of the recognition sequence. The former cut will generate “blunt ends” with no nucleotide overhangs. The latter generates “sticky” or “cohesive” ends, because each resulting fragment of DNA has an overhang that compliments the other fragments. Both are useful in molecular genetics for making

Restriction enzymes are exceedingly varied; they range in size from the diminutive PvuII (157 amino acids) to the giant CjeI (1,250 amino acids) and beyond. Among over 3,000 activities that have been purified and characterized, more than 250 different sequence specificities have been discovered [3]. The search for new specificities continues, both biochemically, by the analysis of cell extracts, and computationally, by the analysis of sequenced genomes. Although most activities encountered today turn out to be duplicates—isoschizomers—of existing specificities, restriction enzymes with new specificities are found with regularity.

Restriction enzymes are used in biotechnology to cut DNA into smaller strands in order to study fragment length differences among individuals (Restriction Fragment Length Polymorphism—RFLP) or for gene cloning. RFLP techniques have been used to determine that individuals or groups of individuals have distinctive differences in gene sequences and restriction cleavage patterns in certain areas of the genome. Knowledge of these unique areas is the basis for DNA fingerprinting. Each of these methods depends on the use of agarose gel electrophoresis for separation of the DNA fragments.

Most DNA in the genome is not involved in coding sequences; much of this DNA is thus not subject to strong

selective pressure for maintaining identical sequences from one individual to the next. Within noncoding DNA regions unrelated individuals exhibit approximately one base pair change per 200 bp. These functionally silent variations are inherited according to strict Mendelian genetics. Given the existence of several 100 restriction enzymes, many such changes can be detected via the appearance or disappearance of a restriction enzyme cleavage site associated with the altered base pair, i.e., the DNA sequence GAATTC is uniquely recognized and cleaved by the restriction enzyme EcoRI; point mutations of the EcoRI sequence, such as CAATTC, AAATTC, TAATTC, GGATTC, GCATTC, and GTATTC, are not cleaved by EcoRI. Thus any of these point mutations within a particular EcoRI site would cause the site to disappear. In rarer instances RFLPs can also occur within coding DNA.

An alternative cause of RFLPs involves no DNA sequence changes within restriction enzyme cleavage sites, but rather, the occurrence of repetitive DNA between two successive restriction sites. (Return to Topic 3 for a quick refresher course in repetitive DNA.) Remember that repetitive DNA regions are very prone to variations in the number of reiterated elements (e.g., ... CACACA... could be repeated 30 times in a particular region of one chromosome and 40 times within the same region of the homologous chromosome; similarly, a 28-bp-long VNTR could occur five times in a particular region of one chromosome and nine times within the same region of the homologous chromosome). When such variable length repetitive DNA sequences occur between two successive restriction enzyme cleavage sites, they cause the total bp number between the two sites to vary. This can be detected by hybridizing a unique probe to this region. Homozygosity for either of the RFLP fragment lengths would be indicated in Southern blots by a single DNA band equivalent to either the “long” or “short” fragment; heterozygosity would be indicated by a DNA band of each size.

Restriction enzymes recognize specific nucleotide sequences within DNA molecules. However, the recognition specificity of restriction enzymes can be reduced *in vitro*. Under certain conditions, enzymes are able to recognize and cleave nucleotide sequences which differ from the canonical site. At low ionic strength, for example, BamHI (with the recognition sequence GGATCC) is able to cleave the following sequences: NGATCC, GPuATCC and GGNTCC. This phenomenon is called “relaxed” or “star” activity [4–6] (Fig. 3.1)

In most practical applications of restriction enzymes, star activity is not desirable. Analysis of several reports on the star activity suggests the following causes for this phenomenon:

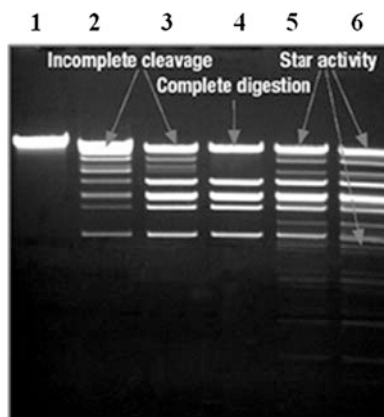


Fig. 3.1 *Enzyme star activity*: 1—Lambda DNA, 2—Lambda DNA incubated 1 h with 0.15 U of EcoRI (incomplete cleavage), 3—Lambda DNA incubated 1 h with 0.4 U of EcoRI (incomplete cleavage), 4—Lambda DNA incubated 1 h with 1 U of EcoRI (complete, digestion), 5—Lambda DNA incubated 16 h with 40 U of EcoRI (star activity), 6—Lambda DNA incubated 16 h with 70 U of EcoRI (star activity)

1. Prolonged incubation
2. High enzyme concentration in the reaction mixture
3. High glycerol concentration in the reaction mixture
4. Presence of organic solvents, such as ethanol or dimethyl sulfoxide, in the reaction mixture
5. Low ionic strength of the reaction buffer
6. Suboptimal pH values of the reaction buffer
7. Substitution of Mg^{2+} for other divalent cations, such as Mn^{2+} or Co^{2+} [7]

In some cases, the termini generated by DNA cleavage with a restriction enzyme at the canonical site have been shown to stimulate the enzyme's star activity. Star activity and incomplete DNA digestion result in atypical electrophoresis patterns, which can be identified by careful examination of gel images (*see* picture below). Here, incomplete DNA digestion results in additional low-intensity bands above the expected DNA bands on the gel. No additional bands below the smallest expected fragment are observed. These additional bands disappear when the incubation time or amount of enzyme is increased. On the contrary, star activity results in additional DNA bands below the expected bands and no additional bands above the largest expected fragment. These additional bands become more intense with the increase of either the incubation time or the amount of enzyme, while the intensity of the expected bands decreases. Some restriction enzymes may remain associated with the substrate DNA after cleavage and thus change the mobility of digestion products

during electrophoresis. The resulting atypical pattern is not related to star activity (Fig. 3.1). To avoid confusing electrophoresis patterns, use a loading dye with SDS (e.g., the Fermentas 6× DNA Loading Dye & SDS Solution). Then, heat the sample for 10 min at 65 °C and place it on ice prior to loading it on the gel. Any tendency of a restriction enzyme to exhibit star activity is indicated both in the product description and in the Certificate of Analysis supplied with each enzyme.

3.2 Materials

1. A 10× stock of the appropriate restriction enzyme buffer.
2. DNA to be digested (see Notes 3 and 4) in either water or TE (10 mM Tris-HCl, pH 8.3, 1 mM EDTA).
3. Bovine serum albumin (BSA) at a concentration of 1 mg/ml.
4. Sterile distilled water.
5. The correct enzyme for the digest.
6. 5× loading buffer: 50 % (v/v) glycerol, 100 mM Na₂EDTA, pH 8.0, 0.125 % (w/v) bromophenol blue (6 pb), 0.125 % (w/v) xylene cyanol.
7. 100 mM Sperm DNA.

3.3 Method

1. Thaw all solutions, with the exception of the enzyme, and then place on ice.
2. Decide on a final volume for the digest, usually between 10 and 50 µl, and then into a sterile.
3. Eppendorf tube, add 1/10 volume of reaction buffer, 1/10 volume of BSA, between 0.5 and 1 µg of the DNA to be digested, and sterile distilled water to the final volume.
4. Take the restriction enzyme stock directly from the -20 °C freezer, and remove the desired units of enzyme with a clean sterile pipette tip. Immediately add the enzyme to the reaction and mix.
5. Incubate the tube at the correct temperature (as per the manufacturer instructions) for ~2–6 h. Genomic DNA can be digested overnight.
6. An aliquot of the reaction usually 10–12 µl may be mixed with a 5× concentrated loading buffer and analyzed by gel electrophoresis.

Table 3.1
Single letter code

R = G or A	K = G or T	B = C, G or T
Y = C or T	S = C or G	D = A, G or T
W = A or T	H = A, C or T	N = G, A, T or C
M = A or C	V = A, C or G	

3.4 Note

A number of restriction enzymes discovered by Fermentas are isoschizomers of commonly used prototype restriction enzymes. Table 3.1 will help you find the appropriate Fermentas enzymes for your experiments.

References

1. Smith HO, Nathans D (1973) A suggested nomenclature for bacterial host modification and restriction systems and then enzymes. *J Mol Biol* 81:19–423
2. Fuchs R, Blakesley R (1983) Guide to the use of Type II restriction endonucleases. *Methods Enzymol* 100:3–38
3. Nasri M, Thomas D (1987) Alteration of the specificity of PvuII restriction endonuclease. *Nucleic Acids Res* 15:7677–7687
4. George J, Chirikjian JG (1982) Sequence-specific endonuclease BamHI: relaxation of sequence recognition. *Proc Natl Acad Sci USA* 79:2432–2436
5. Mayer H (1978) Optimization of the EcoRI* activity of EcoRI endonuclease. *FEBS Lett* 90:341–344
6. McClelland M, Hatnsh J, Nelson M, Patel Y (1988) KGB: a single buffer for all restriction endonucleases. *Nucleic Acids Res* 16:364
7. Hsu M, Berg P (1978) Altering the specificity of restriction endonuclease: effect of replacing Mg²⁺ with Mn²⁺. *Biochemistry* 17:131–138

Genetic Fingerprinting Techniques for Molecular Characterisation of Microbes

Annette Reineke and K. Uma Devi

Abstract

DNA fingerprints are commonly generated for a genetic characterisation of microbial populations or communities. The respective techniques are based either on hybridisation or on polymerase chain reaction (PCR). We present an overview and detailed protocols of the most frequently DNA fingerprinting techniques currently used in microbial ecology, including isolation of respective target sequences, set-ups of PCR reactions, and ways of detecting markers for generating fingerprints.

4.1 Introduction

Genetic or DNA fingerprinting (also called DNA profiling or DNA testing) is a term applied to a range of techniques that are used to show similarities and dissimilarities between samples of DNA from different individuals. Since the sequence of nucleotides in an individual's DNA is as unique as its fingerprint, the term genetic fingerprinting was introduced in the mid-1980s initially in the context of assisting in correct identification of humans especially in forensic investigations. However, DNA fingerprinting was quickly shown to work in all kinds of organisms from mammals to plants, invertebrates, and microorganisms as well.

In microbial ecology, it is often important to have tools at hand for genetic characterisation of a community of interest. Such a characterisation may address all the microbes (the types and their numbers) in the community, or microbes of one taxonomic group (bacteria, fungi, algae, etc.), or microbes with a particular metabolic pathway, e.g. a nitrifying or denitrifying community. Community characterisation can help to answer questions such as has the community changed over time or is the change in community correlated with any environmental parameters? To characterise a microbial community the cliché lies in the fact that only around 0.1–10 % [1] of microbes are cultivable. Even to retrieve the cultivable ones from a

sample requires trying several different culture media. Therefore, molecular techniques are invoked to characterise the community either biochemically, based on the reactions they catalyse, or genetically, based on their DNA sequences using DNA fingerprinting techniques. Such techniques are also applied in cases where the organism of interest is cultivable and the aim of the experiment is to characterise the variability of that species in an ecological niche or to compare these organisms from different regions. The philosophy underlying genetic profiling of microbial communities is that the DNA of all the microbes in the sample is retrieved through the DNA extraction process, the DNA is from living organisms, and there is no bias in the amplification of the DNA sequence selected for any particular genus/species/strain.

DNA fingerprinting techniques are basically of two kinds: hybridisation and PCR (Polymerase Chain Reaction) based. Restriction Fragment Length Polymorphism (RFLP) is a hybridisation-based technique. Among the several PCR-based genetic profiling methods, RAPD (Random Amplified Polymorphic DNA), AFLP (Amplified Fragment Length Polymorphism), SSCP (Single Strand Conformation Polymorphism), and SSR/STR (Simple Sequence/Tandem Repeat) methods are used for studying the population structure of a particular species. The organisms used in these studies are in most cases cultivable.

For analysis of variability in microbial communities, DNA is extracted not from cultured organisms but from a sample such as soil, water, abdominal content of an organism, etc. DNA isolation and purification kits are available for, e.g., soil or water samples. The most commonly used methods are ARDRA (Amplified Ribosomal DNA Restriction Analysis), T-RFLP (Terminal Restriction Fragment Length Polymorphism), and DGGE and TGGE (Denaturing/Temperature Gradient Gel Electrophoresis). T-RFLP method is useful for rapid analysis to assess community diversity, but ARDRA is more discriminative to differentiate phylotypes.

Quite frequently, ribosomal RNA genes are used for genetic profiling of microbial communities. There are several reasons for rRNA genes to be the ones of choice: (a) they occur universally in all organisms, (b) they have long, highly conserved regions useful for looking for distant phylogenetic relationships, (c) they have sufficient variable regions to assess close relationship, (d) they are not prone to rapid sequence change due to selection and serve as evolution chronometer. Due to the extensive use of rRNA genes in genetic profiling of communities, a database has been created including the patterns in different studies (<http://rdp.cmc.msu.edu>). Therefore if rRNA gene profiling is done it can be readily compared with data in the database which would facilitate easy identification of the microbes. In addition, any DNA/gene sequence suitable for the purpose of the experiment also can be targeted for profiling of microbial communities (Table 4.1).

Table 4.1
Examples of universal genes and primer sequences targeted in various genetic profiling methods

Genes targeted	Primer sequence (5'–3')	Reference primer original publication	Example of method	Reference application of the technique
ITS1 region of ribosomal RNA	ITS5: GGAGTAAAAGTCGTAACAAGG	[2]	SSCP	[3]
	ITS2: GCTGCGTTCTTCATCGATGC			
β tubulin	Bt2a: GGTAACCAAAATCGGTGCTGCTTTC	[4]	SSCP	[5]
	Bt2b: ACCCTCAGTGTAGTGACCCCTTGGC			
mt rRNA, small subunit	MS1: CAGCAGTCAAGAATATTAGTCAATG	[2]	SSCP	[5]
	MS2: GCGGATTATCGAAATTAAATAAC			
mt rRNA, large subunit	ML1: GTACTTTTGCATAATGGGTCAGC	[2]	SSCP	[5]
	ML2: TATGTTTCGTAGAAAACCCAGC			
Bacterial 16S rRNA	fD1: AGAGTTTGATCCTGGCTCAG	[6]	ARDRA	[7]
	rD1: AAGGAGGTGATCCAGCC			
Eukaryotic 16S rRNA	EukA: CCGAATTCGTGACAAACCTGGTTGATCCTGCCAGT	[8]	ARDRA	[9]
	EukB: CCGGGGATCCAAAGCTTGATCCTTCTGCAGGTTTACCTAC			
Fungal small subunit rDNA	nu-SSU-0817-5: TTAGCATGGAATAATRAATAGGA	[10]	T-RFLP	[11]
	nu-SSU-1196-3: TCTGGACCTGGTGAGTTTCC			
	nu-SSU-1536-3: ATTGCAATGCYCTATCCCCA			
Universal 18S rDNA primers	530F: GTGCCAGMCGCCGG	[12]	T-RFLP	[13]
	519R: GWATTACCGCGGCKGCTG			
ITS1 region of ribosomal RNA	ITS1: TCCGTAGGTGAACCTGCGG	[2]	T-RFLP	[14]
	ITS4: TCCTCCGCTTATTGATATGC			
Bacterial 16S rRNA, V3 region	PRBA 338F 5' ACTCCTACGGGAGGCAGCAG	[12]	DGGE	[15]
	PRUN 518R 5' ATTACCGCGGCTGCTGG			

The conditions of the PCR cycle would depend on the sequence to be amplified. The choice of the restriction enzymes is again a matter of trial and error. Frequent cutters (four bp recognition sites) are commonly used. When sequence information is available, enzymes with the highest number of recognition sites are identified using software such as NEB cutter (<http://tools.neb.com/NEBcutter2>).

Random Amplified Polymorphic DNA (RAPD) fingerprinting is a low-cost yet powerful molecular typing method for organisms from bacterial species to mammals [17, 18]. RAPDs have become well-established genetic tools for genomic mapping and linkage analysis, genotype fingerprinting and identification, and quantification of genetic relationships, similarities, and variation in a variety of organisms. RAPDs are generated from genomic DNA by PCR amplification of only a small amount of total DNA using short (usually 10-mer) randomly constructed oligonucleotides. The advantages of this technique are the essentially unlimited number of loci that can be examined, no need for prior DNA sequence knowledge, and low amount of template DNA required. However, reported limitations of RAPD markers include a low reproducibility of banding patterns since the quality and concentration of template DNA, concentrations of PCR components, presence of inhibitors, or PCR cycling conditions may influence the results obtained [19–21]. Thus, RAPDs require a careful experimental set-up and careful laboratory habits. Furthermore, nearly all RAPD markers behave as dominant markers; therefore, it is not possible to distinguish whether a specific DNA fragment amplified from a locus is heterozygous (present in one copy) or homozygous (present in two copies) in the organism analysed.

Amplified Fragment Length Polymorphism (AFLP) analysis is based on the restriction digestion of genomic DNA, followed by ligation of adapters specific to the restriction sites, and subsequent PCR amplification with adapter-specific primers [22]. This multi-locus DNA profiling technique has successfully been used for identifying polymorphisms in both prokaryotic and eukaryotic organisms and allows the reliable identification of a high number of loci in a single assay [23–25]. For AFLP analysis, only tiny amounts of purified genomic DNA are needed; it does not require prior knowledge of DNA sequences, detects variation over the entire genome, and has proved to be robust and reliable because it uses stringent reaction conditions. As a consequence, it has become one of the most frequently used methods over the last 15 years for linkage mapping, analysis of genetic diversity, population genetics, and single-locus PCR marker development in a wide variety of organisms such as bacteria, fungi, insects, plants, and animals. Genetic variation detected using the AFLP method is represented by the presence or absence of amplified DNA fragments (bands), which are classified as dominant markers. The

presence of a band implies that the locus is either homozygotic (AA: both haploids produce homologous fragments) or heterozygotic (Aa: one haploid produces a fragment), resulting in the inability to distinguish heterozygotes from homozygotes in AFLP analysis. However, the high number of polymorphic loci detected in AFLP analysis can compensate for this shortcoming because the number of loci is a crucial factor for estimating reliable population genetic parameters.

Regions within DNA sequences where short sequences of nucleotides are repeated one after the other (in tandem arrays) are referred to as “microsatellites” [26]. The repeat is usually a short sequence consisting of two, three, or four nucleotides (referred to as di-, tri-, or tetranucleotide repeats, respectively). Therefore, they are referred to as Simple Sequence Repeats (SSRs) or Short Tandem Repeats (STR). The number of repeats at a particular location within a genome can vary between individuals within the same species and therefore they are also tagged as Variable Number Tandem Repeats (VNTR). For example, in some individuals the repeated unit may occur 6 times; in others it may be 12, or 3, or 50. Quite frequently, the number of repeats and the type of repeat are designated in a formula such as (GT)₁₂, where GT refers to the specific nucleotides that are repeated and the subscript (12) designates the number of times the sequence is repeated. In diploid organisms, each individual will have two copies of a particular microsatellite segment, so each locus has its own set of microsatellite alleles. Accordingly, alleles at a specific locus can differ in the number of repeats, e.g. an individual may have eight microsatellite repeats in one allele and—in the case of heterozygous individuals—ten repeats in the other allele. Polymorphic microsatellite loci are ideal molecular markers, which have a wide range of applications including the determination of paternity, population genetic studies, and recombination mapping. Since microsatellite repeats tend to occur in noncoding regions of the DNA, they are selectively neutral. They are also usually inherited in a co-dominant Mendelian fashion.

Single Strand Conformation Polymorphism (SSCP) analysis takes advantage of the fact that slight differences in the nucleotide sequence of two single-stranded DNA fragments can result in a different secondary structure with dissimilar migration speed during electrophoretic separation in a gel matrix [27–29]. Therefore, it is possible to detect polymorphisms of even a single base pair between two DNA samples. SSCP analysis offers an inexpensive, convenient, and sensitive method for determining genetic variation at single or multiple locations in DNA fragments and is frequently used both for generating genetic markers in population genetic studies as well as a mutation scanning technique for diagnostic purposes. It can also be used for distinguishing homozygous and heterozygous states of two alleles, since each alternative

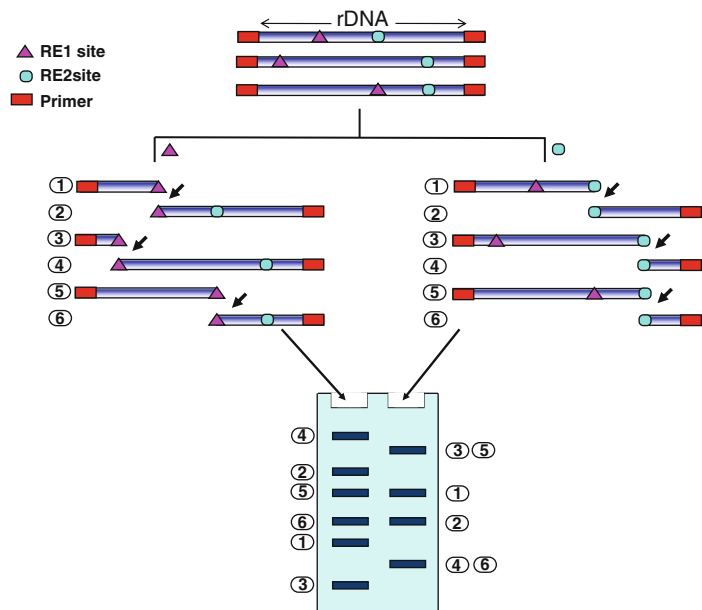


Fig. 4.1 Schematic outline of ARDRA. DNA is amplified with primers specific for the rRNA gene cluster (the 16S or 18S); amplicons are digested with restriction enzymes to reflect the slight differences in the sequences of the different DNA samples.

should result in a different conformation, and thus, in distinct patterns after electrophoresis. However, problems with reproducibility of the SSCP banding pattern have been reported, since changes in concentrations of $MgCl_2$, DNA template, and PCR primers in the PCR reaction, as well as the choice and concentrations of chemical denaturants used before electrophoretic separation, may significantly influence SSCP pattern formation [30].

For profiling of microbial communities, Amplified Ribosomal DNA Restriction Analysis (ARDRA) is a very simple method frequently used, which is based on the analysis of restriction enzyme digestion of the amplified 16SrRNA genes of bacterial species [31, 32]. For eukaryotic microbes, alternatively the 18S rRNA genes can be targeted (Fig. 4.1). When universal 16S rRNA gene primers are used, the analysis gives little or no information on the type of microorganisms in the sample but can be used to quickly assess genotypic changes in a community over time or to compare communities subjected to different environmental conditions. If primers specific for a particular genus are designed and used, this technique can be employed for molecular phylogenetic studies as well. Species-specific primers can be used to find markers for strain identification (Table 4.1).

Terminal restriction fragment length polymorphism (T-RFLP) allows the fingerprinting of a microbial community by analysing the polymorphism of a gene. The method was first described by [33]

with the amplification of the 16S rDNA gene. But any suitable gene can be chosen (Table 4.1). It is a high-throughput, reproducible method that allows semiquantitative analysis of the diversity of a particular gene in a community [14].

Denaturing (DGGE) or Temperature Gradient Gel Electrophoresis (TGGE) are related techniques in which allelic differences in DNA sequences of interest are detected after PCR amplification based on their difference in migration on a gel with a denaturing gradient [34, 35]. The denaturing gradient is created either through difference in the chemical constitution of the gel with an increasing gradient of denaturant (urea and formaldehyde) (DGGE) or through differences in temperature along the length of the gel (TGGE). These techniques are related in principle to the SSCP technique. The difference lies in that, in SSCP, migration differences in allelic forms of a DNA sequence are detected through difference in the conformational modes taken up by single-stranded DNA, while in D/TGGE, the detection is based on mobility of double-stranded DNA of allelic forms of a DNA sequence. The denaturation point is species specific.

RFLP is a standard hybridisation-based DNA fingerprinting technique; in fact it is the first method to have been devised to analyse genetic diversity in a population of a species [36]. It can also serve as an identification mark (barcode) for the individual. It involves restriction digestion of DNA, Southern blotting, and hybridisation with a probe sequence. The restriction enzyme and probe combinations are decided based on the aim of the experiment. A single locus probe is used to analyse polymorphism in a single gene (a typical RFLP) or a multilocus probe is used to analyse the overall DNA polymorphism (a DNA fingerprint) in the sample.

This chapter presents some of the most frequently used methods for genetic fingerprinting and profiling of microbial communities including a selection of suitable protocols for isolation and application of the given molecular marker. For all the fingerprinting techniques presented, we end up with a gel or autoradiogram showing bands or an electropherogram showing peaks. The gel/autoradiogram picture is captured in a digital form and the banding pattern is analysed using suitable software.

4.2 Requirements

4.2.1. Common Requirements for All Protocols

1. Thermocycler
2. 0.2 ml PCR tubes or 96-well microplates
3. DNA in water or TE buffer

4. dNTP mix
5. 10× PCR buffer (usually provided with the respective polymerase)
6. *Taq* DNA polymerase
7. Primers [different for each technique and dependent on the aim of the experiment in some cases (Table 4.1)]
8. PCR product or gel purification kit (for some applications)
9. Agarose/polyacrylamide or premade polyacrylamide solutions
10. 1× TBE buffer: 0.1 M Tris-HCl, 0.1 M boric acid, 2 mM EDTA, pH 8.0
11. 1× TAE buffer: 40 mM Tris, 40 mM acetic acid, 1 mM EDTA, pH 7.6 with glacial CH₃COOH or Tris
12. 1× TE buffer: 10 mM Tris-HCl, 1 mM EDTA, pH 8.0
13. 6× loading dye: 15 % (w/v) Ficoll 400, 0.25 % bromophenol blue, 1× TAE
14. DNA size marker (in the size range of the DNA fragments expected)
15. DNA staining dye like ethidium bromide/SYBR Green

4.2.2. Random Amplified Polymorphic DNA

1. RAPD primer

4.2.3. Amplified Fragment Length Polymorphism

1. Restriction enzymes, such as *Eco*RI (6-bp cutter) and *Mse*I (4-bp cutter) with appropriate reaction buffers (the primer and adapter sequences given below will vary depending on the enzymes chosen). *Eco*RI and *Mse*I are frequently used and are given as examples here.
2. T4 DNA ligase with appropriate reaction buffer.
3. Adapters and primers as follows:
 - (a) *Eco*RI adapter forward: 5'-CTCGTAGACTGCGTACC-3'
 - (b) *Eco*RI adapter reverse: 5'-AATTGGTACGCAGTCTAC-3'
 - (c) *Mse*I adapter forward: 5'-GACGATGAGTCCTGAG-3'
 - (d) *Mse*I adapter reverse: 5'-TACTCAGGACTCAT-3'
 - (e) *Eco*-0: 5'-GACTGCGTACCAATTC-3'
 - (f) *Mse*-0: 5'-GATGAGTCCTGAGTAA-3'
 - (g) *Eco*-NN: *5'-GACTGCGTACCAATTCNN-3'
 - (h) *Mse*-NNN: 5'-GATGAGTCCTGAGTAANN-3'

N in the primer sequence designates A, T, C, or G and can be modified in numbers from one to three selective nucleotides at each 3'-end of the primer. The asterisk (*) refers to the fact that usually the *Eco*-selective AFLP

primer is fluorescently labelled in case an automated sequencer is used for electrophoretic separation.

4. 2× Stop solution: 95 % formamide, 10 mM NaOH, 0.25 % bromophenol blue, filter sterilised through a 0.45 µm filter.
5. Reagents for electrophoretic separation as required for the separation/detection system used. Commercially available precast gels can also be used.

4.2.4. SSR and STR (Short Tandem Repeats and Simple Sequence Repeats)

1. Restriction enzyme *RsaI* and its suitable buffer
2. Linker 1: 5'-GTTTAGCCTTGTAGCAGAAGC-3'
3. Linker 2: 5'-pGCTTCTGCTACAAGGCTAAACAAAA-3'
(p indicates phosphorylation)
4. 10 % SDS
5. 20× SSC buffer: 3 M NaCl, 300 mM trisodium citrate, pH 7.0 adjusted with HCl
6. Biotinylated repeat oligonucleotides such as b(GA)₁₅ or b(CA)₁₅
7. Streptavidin-coated magnetic beads (e.g. Dynabeads from Dynal), washed according to the manufacturer's instructions
8. TA cloning vector kit
9. Competent *Escherichia coli* cells
10. One forward and one reverse primer for amplification of the specific target sequence

4.2.5. SSCP (Single Strand Conformation Polymorphism)

1. One forward and one reverse primer for amplification of the specific target sequence (see Table 4.1)
2. 1× formamide dye: 98 % formamide, 10 mM EDTA, 0.025 % bromophenol blue, 0.025 % xylene cyanol FF, pH 8.0

4.2.6. ARDRA (Amplified Ribosomal DNA Restriction Analysis)

1. One forward and one reverse primer for amplification of the specific target sequence
2. Commercial kit for purification of PCR fragments
3. Specific restriction enzyme and its suitable buffer

4.2.7. T-RFLP (Terminal Restriction Fragment Length Polymorphism)

1. Fluorescently labelled primers (see Note 5). If both primers used are labelled, a different dye is used for each. The amplification efficiency of labelled primers tends to be lower than that of unlabelled primers, frequently leading to lower yields. It is necessary to pool several PCR reactions to obtain enough products for further steps (200–300 ng of DNA recommended per restriction digest). Therefore four replicate

50 µl PCR reactions are made for each sample and the amplicons are pooled up.

2. Commercial kit for purification of PCR fragments.

4.2.8. DGGE and TGGE (Denaturing/ Temperature Gradient Gel Electrophoresis)

1. A forward primer with a GC clamp and a reverse primer for amplification of the specific target sequence. If nested PCR is used, the forward primer of the second PCR cycle is designed with the GC clamp. A GC clamp typically has a sequence like 5' CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG G 3'.

4.2.9. RFLP (Restriction Fragment Length Polymorphism)

1. Genomic DNA from individual samples
2. Restriction enzymes
3. DNA size marker
4. Probe sequence
5. Probe labelling kit
6. Blotting Membrane
7. 20× SSC buffer: 3 M NaCl, 300 mM trisodium citrate, pH 7.0 adjusted with HCl
8. Hybridisation solution: 1 % SDS, 1 M NaCl, 10 % Dextran Sulphate, made to 1 l with 50 mM Tris-Cl pH 7.5. The solution can be stored in a freezer and resuspended in a water bath at 65 °C before use
9. 0.1 % SDS
10. Salmon sperm DNA (10 µg/µl)
11. X-ray film
12. X-ray developing and fixing solutions

4.3 Methods

4.3.1. RAPD (Random Amplified Polymorphic DNA)

For generating RAPD markers, an extract of total genomic DNA is amplified with a single short (10-mer) primer of arbitrary nucleotide sequence. These short primers anneal at random sites at a number of locations in the genome. For successful amplification to occur, the binding of the primer should have happened to sequences matching the primer occurring as inverted repeats within a distance between 50 and 3,000 bp. The higher the instances of sequences occurring as inverted repeats within a distance of 3,000 bp (the threshold length for PCR amplification)

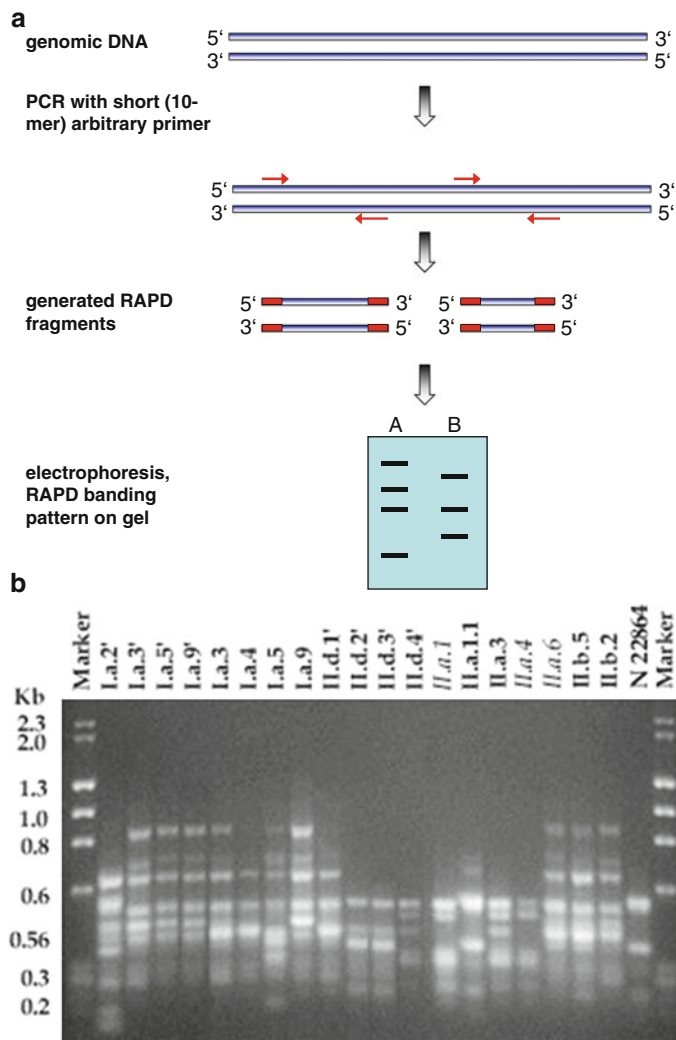


Fig. 4.2 (a) Schematic outline of RAPD analysis. Genomic DNA is amplified using a single short (decamer) primer of arbitrary (random) sequence at a low annealing temperature and amplified products are separated on an agarose gel. (b) Example of banding pattern obtained after RAPD analysis of isolates of the entomopathogenic fungi *Beauveria bassiana* and *Nomuraea rileyi*. M designates a molecular weight marker (picture courtesy of authors).

matching the RAPD primer in the genome of the respective organism, the more amplification products are generated during PCR (Fig. 4.2a). An example picture of a RAPD fingerprint is given in Fig. 4.2b.

When starting an RAPD project, usually several factors have to be carefully adjusted. For instance, concentrations of reaction components such as $MgCl_2$, primer, and dNTPs as well as the quality and concentration of the target DNA have to be experimentally tested for amplification efficiency. Once these parameters are adapted for the organism under study, PCR reactions can be set up. In RAPD, only one primer is used (unlike the traditional

two for standard PCR reactions). RAPD PCR reactions are performed at very low annealing temperature and a high number of PCR cycles—usually 45 cycles are applied. The low annealing temperature (usually 36 °C) is used to permit promiscuous pairing of the primer allowing also single mismatches. This is a trade-off to achieve a reasonable number of amplicons for assessing polymorphism. At the same time, this feature of RAPD analysis is the chief reason for reported problems in reproducibility of the reactions. It has therefore been shown that the number of amplification products in a RAPD reaction is affected by the accuracy of the PCR machine and the fidelity of the *Taq* DNA polymerase used.

RAPD amplification products are separated on a 1–2 % agarose gel, stained with ethidium bromide, silver nitrate, or other suitable dyes, and photographed. Variability between the individuals is scored as the presence (1) or absence (0) of a specific band (amplification product) and various coefficients of genetic diversity can be calculated subsequently.

1. Add the following to the tubes on ice (total volume of 25 µl) (see Notes 1, 2, 3):
 - (a) 17.8 µl sterile distilled H₂O
 - (b) 2.5 µl 10× PCR buffer (contains 15 mM MgCl₂)
 - (c) 1.0 µl RAPD primer (5 pmol/µl)
 - (d) 2.5 µl 2 mM dNTPs
 - (e) 0.2 µl *Taq* DNA polymerase (5 U/µl)
 - (f) 1.0 µl template DNA (diluted to ca. 10 ng/µl) (see Note 4)
2. Mix gently, centrifuge briefly, and place in thermocycler.
3. PCR program
 - (a) Initial denaturation at 94 °C for 2 min
 - (b) Amplification at 45 cycles of
 - 94 °C for 1 min
 - 36 °C for 1 min
 - 72 °C for 2 min
 - (c) Final elongation at 72 °C for 4 min
 - (d) Hold reactions at 4 °C
 - (e) Final elongation at 72 °C for 4 min
 - (f) Hold reactions at 4 °C
4. Electrophoretic Separation
 - Add 3.0 µl 6× loading dye to the tubes on ice, mix, and centrifuge briefly.

- Load 15 µl of the reaction on a 1–2 % agarose gel in 1× TBE or 1× TAE buffer and separate fragments at 80 V for ca. 1 h depending on expected size of the bands.

4.3.2. AFLP (Amplified Fragment Length Polymorphism)

For AFLP fingerprinting, genomic DNA is generally digested with two restriction enzymes, one with an average cutting frequency (e.g. *EcoRI* or other 6 bp-cutters) and the other with a higher cutting frequency (e.g. *MseI* or other 4 bp-cutters). The restriction enzymes are chosen based on the genome size and complexity of the species being studied. After restriction digestion, double-stranded oligonucleotide adapters corresponding to the restriction sites are ligated to the fragments. These adapters serve as binding sites for adapter-specific primers in two subsequent PCR amplifications with an aliquot of the ligation reactions as a template. In the first round of amplification—termed pre-amplification—only those restriction fragments carrying adapters specific for the two restriction enzymes at their 5'- and 3'-end are amplified. In the second round of amplification, basically the same primers are used, differing from the pre-amplification primers by having an extension of one to three selective nucleotides at their 3' end. This allows the amplification of only a subset of fragments from the initial pool of pre-amplified PCR fragments under highly stringent conditions. If both primers have, e.g., two selective nucleotides at their 3'-ends, 16 primers per site are possible with a combination of 256 different primers amplifying the whole population of possible fragments. The strategy behind this approach is to reduce the number of amplicons and therefore bands to a reasonable number to make downstream analysis more practical (Fig. 4.3a).

After amplification, AFLP fragments are loaded on a denaturing polyacrylamide gel or another gel matrix with a high-resolution capacity. Initially, AFLP fragments were radioactively labelled either by using radioactive end labelling of one of the AFLP primers or in the course of PCR amplification via incorporation labelling of radioactive nucleotides. These days, one of the AFLP primer used during selective amplification is commonly labelled with a fluorophore and fluorescently labelled fragments are subsequently detected during electrophoretic separation on an automated sequencer. Using several different AFLP primers each labelled with a different fluorophore during selective amplification allows the set-up of a multiplex PCR reaction, where AFLP products are separately detected for each fluorophore. An example of an AFLP fingerprint based on fluorescently labelled fragments is shown in Fig. 4.3b. Alternatively, AFLP fragments are not labelled during PCR, but gels are stained after electrophoresis with silver nitrate or any other sensitive DNA staining method like SYBR Safe or ethidium bromide.

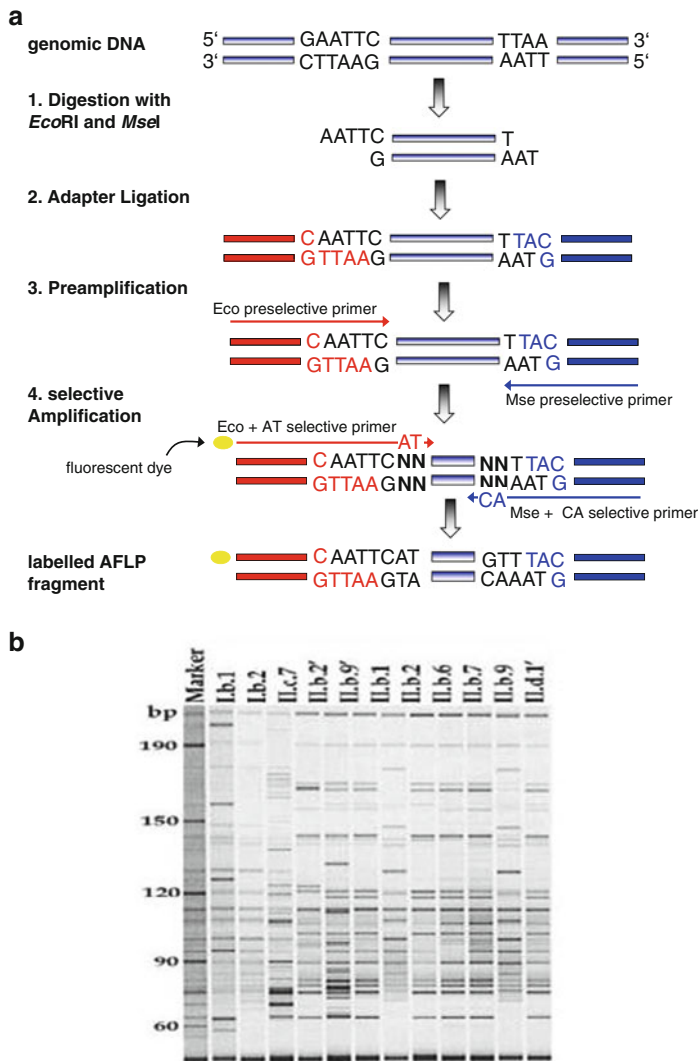


Fig. 4.3 (a) Schematic outline of AFLP analysis. Genomic DNA is digested with two restriction enzymes; adapters are ligated to the restriction fragments which serve as primer binding sites for subsequent PCR reactions. During selective amplification, primers matching the adaptor sequence with additional one, two, or three nucleotides are used. (These primers are labelled with a fluorescent tag if the electrophoresis system used has a fluorescence detection system such as most capillary DNA sequencers.) The amplified products are separated on polyacrylamide gels or via capillary electrophoresis. (b) Example of part of a banding pattern obtained after AFLP analysis of 11 isolates of the entomopathogenic fungus *Nomuraea rileyi*. M designates a molecular weight marker.

1. Pipette together:
 - (a) $\leq 9 \mu\text{l}$ DNA (200 ng in $\leq 9 \mu\text{l}$) (see Note 4)
 - (b) $1.25 \mu\text{l}$ $10\times$ appropriate restriction enzyme buffer (with $100 \mu\text{g/ml}$ BSA)
 - (c) $0.2 \mu\text{l}$ *Eco*RI ($20 \text{ U}/\mu\text{l} = 5 \text{ U}$, NEB)
 - (d) $0.3 \mu\text{l}$ *Mse*I ($10 \text{ U}/\mu\text{l} = 3 \text{ U}$, NEB)

- (e) Fill up to a total volume of 12.5 μl with sterile distilled H_2O
2. Mix gently, centrifuge briefly, and incubate at 37 °C for 2 h.
3. Inactivate restriction enzymes by incubating 15 min at 65 °C, and place on ice (or hold at 4 °C).
4. Preparation of the adapter solutions
 - (a) Add forward and reverse adapters for *Eco*RI and *Mse*I, in separate tubes from a stock solution of 100 pmol/ μl in equal volume to one tube stock and mix them.
 - (b) Heat at 65 °C for 10 min, and allow to cool at room temperature for 1–2 h.
5. Add the following to the tubes (on ice):
 - (a) 12.5 μl digested DNA (complete digest)
 - (b) 2.5 μl 10 \times T4 DNA Ligase buffer
 - (c) 1.0 μl *Eco*RI-adapter solution (5 pmol/ μl)
 - (d) 1.0 μl *Mse*I-adapter solution (50 pmol/ μl)
 - (e) 0.5 μl T4 DNA Ligase (400 U/ μl = 200 U)
 - (f) 7.5 μl sterile distilled H_2O
6. Mix gently, centrifuge briefly, and incubate at 16 °C for 2 h, and inactivate enzyme by incubating 15 min at 65 °C.
7. Perform a 1:10 dilution of the ligation mixture by transferring 10 μl of the ligation reaction to a new tube/plate and add 90 μl TE buffer, and freeze samples. Store the unused proportion (15 μl) of the ligation reaction at –20 °C for long-term use.
8. Add the following to the tubes on ice (total volume of 10 μl) (see Notes 1, 2, 3):
 - (a) 3.8 μl sterile distilled H_2O
 - (b) 1.0 μl 10 \times PCR buffer (contains 15 mM MgCl_2)
 - (c) 1.0 μl primer *Eco*-0 (5 pmol/ μl)
 - (d) 1.0 μl primer *Mse*-0 (5 pmol/ μl)
 - (e) 1.0 μl 2 mM dNTPs
 - (f) 0.2 μl *Taq* DNA polymerase (5 U/ μl)
 - (g) 2.0 μl diluted ligation reaction mixture
9. Mix gently, centrifuge briefly, and place in thermocycler.
10. PCR program:
 - (a) Amplification at 20 cycles of
 - 94 °C for 30 s
 - 56 °C for 1 min

- 72 °C for 1 min
- (b) Hold reactions at 4 °C
- 11. Dilute an aliquot of the reaction 1:50 with sterile distilled H₂O.
- 12. Store reactions and diluted aliquots at −20 °C.
- 13. Add the following to the tubes on ice (total volume of 11 µl) (see Notes 1, 2, 3):
 - (a) 5.1 µl sterile distilled H₂O
 - (b) 1.1 µl 10× PCR buffer (contains 15 mM MgCl₂)
 - (c) 1.0 µl primer *Mse* primer + N (5 pmol/µl)
 - (d) 0.5 µl fluorescently labelled *Eco*RI primer + NN (1 pmol/µl) (see Note 5)
 - (e) 1.1 µl 2 mM dNTPs
 - (f) 0.2 µl *Taq* DNA polymerase (5 U/µl)
 - (g) 2.0 µl diluted pre-amplified DNA
- 14. Mix gently, centrifuge briefly, and place in thermocycler.
- 15. PCR program:
 - (a) Amplification at 12 cycles of
 - 94 °C for 30 s
 - 65 °C for 30 s, with a decrease in annealing temperature of 0.7 °C/cycle during the following cycles
 - 72 °C for 1 min
 - (b) Amplification at 23 cycles of
 - 94 °C for 30 s
 - 56 °C for 1 min
 - 72 °C for 1 min
 - (c) Hold reactions at 4 °C
- 16. Store reactions at 4 °C or −20 °C before electrophoretic separation.
- 17. Add 5.0 µl 2× Stop Solution to the tubes on ice, mix, and centrifuge briefly.
- 18. Denature samples for 3 min at 94 °C and then quickly cool on ice.
- 19. Load 1.0 µl (or more, depending on gel type) on a polyacrylamide gel and separate fragments according to the specifications of the electrophoretic apparatus to be used.
- 20. Analyse presence/absence of bands according to the technique and software used.

4.3.3. SSR and STR (Short Tandem Repeats and Simple Sequence Repeats)

Commonly, microsatellite loci are detected in individuals by PCR, using primers that bind to unique sequences of the SSR flanking regions. A single pair of PCR primers will therefore be useful for each SSR in a given species and produce different sized products depending on the number of microsatellite repeating units in each allele. It is possible in this technique to differentiate between homozygous and heterozygous condition of an SSR locus. The microsatellite primers are usually developed by inserting random segments of the genomic DNA of interest into a vector and cloning them in bacteria (Fig. 4.4a). Bacterial colonies containing these segments are screened with labelled oligonucleotides that hybridise to a given microsatellite repeat. Positive clones are sequenced and PCR primers are designed based on the SSR flanking sequences. The primers are subsequently used in standard PCR reactions (Fig. 4.4b) and the resulting PCR products are separated by either electrophoresis on gels with high-resolution capacity or capillary electrophoresis. The size of the PCR product(s) allows the determination of how many times the given short sequences of nucleotides is repeated for each allele. Sometimes not only the expected one or two major bands are produced during microsatellite PCR but often there are additional minor bands visible. These bands are called stutter bands differing from the major bands by a few nucleotides and are a result of misamplification of the locus during the PCR process. An example of an SSR fingerprint is represented in Fig. 4.4c.

Conventionally, microsatellite loci are identified in a species of interest as described above using genomic libraries selected for small insert size and screening several thousands of clones through colony hybridisation with repeat containing probes. Although this is a relatively simple approach, this process of identification can become really tedious especially when working with microsatellite rich genomes, while ineffective, for species with low microsatellite frequencies. Thus, several enhanced protocols have been published which aim at increasing the proportion of genomic DNA fragments containing repeat motifs in microsatellite-enriched libraries [37]. In most of the cases, such an enrichment is accomplished by hybridising linker-ligated genomic DNA to synthesised oligonucleotide repeats, which are labelled, e.g., with biotin. As biotin binds strongly to streptavidin, if streptavidin-coated iron beads are subsequently used for hybridisation, the biotinylated oligos, along with bound strands of genomic DNA, will bind to the beads. After removing any genomic DNA not bound to these oligos, a repeat-enriched genomic DNA solution is obtained and is used for further cloning and analysis. The following protocol includes an enrichment procedure and is based on the methods published by [38, 39].

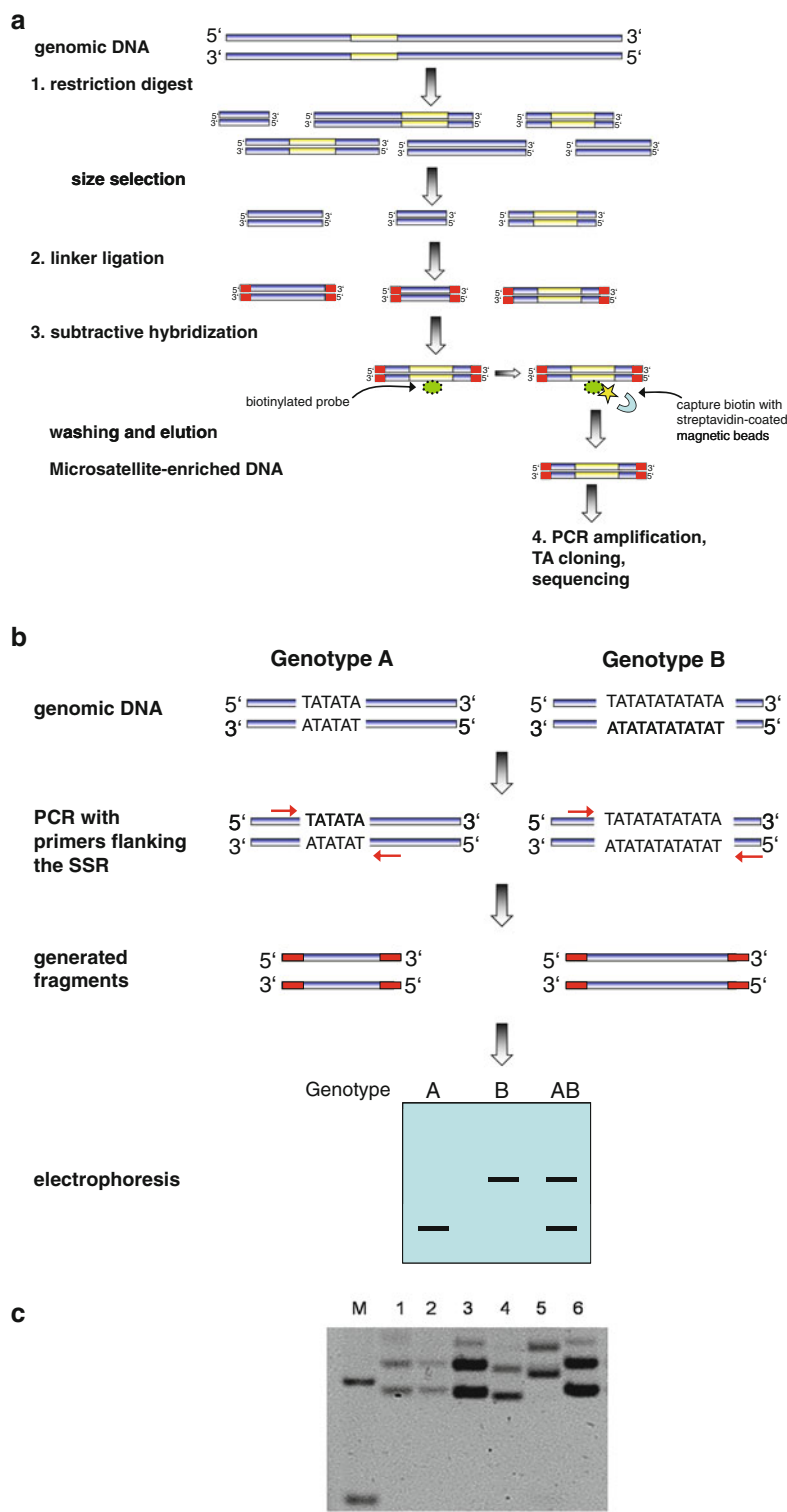


Fig. 4.4 (a) Steps during isolation of SSR markers and generation of an SSR-enriched library. (b) Schematic outline of SSR analysis. Genomic DNA is amplified with primers flanking the previously identified SSRs and obtained fragments are separated via electrophoresis. (c) Example of a banding pattern obtained after SSR analysis of six individuals of a lepidopteran insect, *Lobesia botrana*. M designates a molecular weight marker (picture courtesy of authors).

1. Pipette together:
 - (a) ≤ 44 μl genomic DNA (4–10 μg DNA in ≤ 44 μl)
 - (b) 5.0 μl *Rsa*I 10 \times restriction buffer
 - (c) 2.0 μl *Rsa*I (20 U/ μl)
 - (d) Fill up to a total volume of 50 μl with sterile distilled H_2O
2. Mix gently, centrifuge briefly, and incubate at 37 °C for 2 h or longer.
3. Inactivate restriction enzymes by incubating 15 min at 65 °C, and place on ice (or hold at 4 °C).
4. Separate fragments on a 1 % agarose gel, and gel purify fragments between ca. 300 and 1,000 bp using a commercial gel purification kit according to the manufacturer's instructions.
5. Preparation of double-stranded linker solutions
 - (a) Add each 10 μM forward and reverse linkers (linker 1 and linker 2) in equal volumes to one tube stock and mix them; this makes a 5 μM double-stranded linker stock solution.
 - (b) Heat at 65 °C for 10 min, and allow to cool at room temperature for 1–2 h.
6. Add the following to the tubes on ice (total volume of 30 μl):
 - (a) 10.0 μl digested, size-selected, and gel-purified DNA (2–5 μg)
 - (b) 3.0 μl 10 \times T4 DNA Ligase buffer
 - (c) 10.0 μl double-stranded linker solution (5 μM)
 - (d) 2.0 μl T4 DNA Ligase (400 U/ μl)
 - (e) 1.0 μl restriction enzyme *Xmn*I (20 U/ μl)*
 - (f) 4.0 μl sterile distilled H_2O

*Note: During the ligation/restriction reaction, both linkers will form an *Xmn*I restriction site. Including *Xmn*I in the ligation reaction will cut these dimers apart, which will maintain a large pool of monomer linkers for ligation to genomic DNA fragments.
7. Mix gently, centrifuge briefly, and incubate at 16 °C for 2 h; inactivate enzymes by incubating 15 min at 65 °C.
8. To ensure that the linkers had ligated to the genomic DNA, perform a PCR with the ligated DNA:
9. Add the following to the tubes on ice (total volume of 20 μl):
 - (a) 12.8 μl sterile distilled H_2O

- (b) 2.0 μ l 10 \times PCR buffer (contains 15 mM MgCl₂)
 - (c) 1.0 μ l linker 1 (0.8 μ M final concentration)
 - (d) 2.0 μ l 2 mM dNTPs
 - (e) 0.2 μ l *Taq* DNA polymerase (5 U/ μ l)
 - (f) 2.0 μ l template DNA (ligated genomic DNA)
10. Mix gently, centrifuge briefly, and place in thermocycler.
11. PCR program:
- (a) Initial denaturation at 94 °C for 2 min
 - (b) Amplification at 30 cycles of
 - 94 °C for 30 s
 - 60 °C for 1 min
 - 68 °C for 1 min
 - (c) Final elongation at 68 °C for 5 min
12. Run products on 1 % agarose gel; a smear is expected within the size range of 200–1,000 bp if linkers have successfully ligated.
13. Pipette together for a total volume of 200 μ l:
- (a) 2.0 μ l linker-ligated genomic DNA (50 ng/ μ l)
 - (b) 100.0 μ l 2 \times hybridisation solution (12 \times SSC, 0.1 % SDS)
 - (c) 2.0 μ l biotinylated repeat oligonucleotide (20 nM)
 - (d) Fill up to a total volume of 200 μ l with sterile distilled H₂O
14. Mix gently, centrifuge briefly, denature for 15 min at 95 °C, and incubate at 60 °C (or other hybridisation temperature, depending on the T_m of the repeat oligo) overnight.
15. Mix with 600 μ g of streptavidin-coated magnetic beads.
16. Incubate at 43 °C for 2 h with continuous agitation.
17. Wash beads twice in 2 \times SSC, 0.1 % SDS at room temperature for 5 min*.
18. Wash beads twice in 1 \times SSC, 0.1 % SDS at 60 °C for 5 min*.
- *Note: Wash temperatures can be adjusted to increase (hotter) or decrease (cooler) hybridisation stringency.
19. Elute repeat-enriched genomic DNA from the beads with 60 μ l preheated TE buffer by incubating at 95 °C for 10 min and immediately recovering the eluate.
20. Add the following to the tubes on ice (total volume of 50 μ l) (see Notes 1, 2, 3):
- (a) 12.8 μ l sterile distilled H₂O

- (b) 5.0 μl 10 \times PCR buffer (contains 15 mM MgCl_2)
 - (c) 1.0 μl Linker 1 (0.8 μM final concentration)
 - (d) 2.0 μl 2 mM dNTPs
 - (e) 0.2 μl *Taq* DNA polymerase (5 U/ μl)
 - (f) 5.0 μl template DNA (repeat-enriched DNA)
21. Mix gently, centrifuge briefly, and place in thermocycler.
22. PCR program:
- (a) Initial denaturation at 94 °C for 2 min
 - (b) Amplification at 25–30 cycles of
 - 94 °C for 30s
 - 60 °C for 1 min
 - 68 °C for 1 min
 - (c) Final elongation at 68 °C for 7 min
23. Run a 5 μl aliquot of the amplified enriched DNA on a 1 % agarose gel to check amplification.
24. Purify the PCR reaction with a commercial PCR purification kit.
25. Adenylate the 3'-ends of the PCR products by adding the following to the tubes on ice (total volume of 5 μl):
- (a) 1.5 μl sterile distilled H_2O
 - (b) 0.5 μl 10 \times PCR buffer (contains 15 mM MgCl_2)
 - (c) 1.0 μl 2 mM dATP
 - (d) 1.0 μl *Taq* DNA polymerase (5 U/ μl)
 - (e) 1.0 μl purified PCR products (150–200 ng/ μl)
26. Mix gently, centrifuge briefly, place in thermocycler, and incubate at 70 °C for 30 min.
27. Ligate adenylated PCR products into a TA cloning vector according to the instructions provided by the supplier and using standard procedures for transforming competent cells and isolation of plasmid DNA.
28. Sequence plasmid DNA from positive recombinant clones using standard sequencing procedures.
29. Design primers based on the sequences flanking the microsatellite arrays identified within the sequenced inserts.
30. Primers must be tested and optimised in order to ensure faithful and consistent amplification.
31. Add the following to the tubes on ice (total volume of 15 μl) (see Notes 1, 2, 3):

- (a) 8.9 μl sterile distilled H_2O
 - (b) 1.5 μl 10 \times PCR buffer (contains 15 mM MgCl_2)
 - (c) 1.0 μl fluorescently labelled forward primer (4 pmol/ μl) (see Note 5)
 - (d) 1.0 μl reverse primer (10 pmol/ μl)
 - (e) 1.5 μl 2 mM dNTPs
 - (f) 0.1 μl *Taq* DNA polymerase (5 U/ μl)
 - (g) 1.0 μl template DNA (diluted to ca. 10–50 ng/ μl) (see Note 4)
32. Mix gently, centrifuge briefly, and place in thermocycler.
33. PCR program:
- (a) Initial denaturation at 94 °C for 2 min
 - (b) Amplification at 20 cycles of
 - 94 °C for 30 s
 - 65 °C for 30 s, with a decrease in annealing temperature of 0.5 °C/cycle during the following cycles
 - 72 °C for 30 s
 - (c) Amplification at 15 cycles of
 - 94 °C for 30 s
 - 55 °C for 30 s (see Note 6)
 - 72 °C for 30 s
 - (d) Final elongation at 72 °C for 3 min
 - (e) Hold reactions at 4 °C
34. Store reactions at $-20\text{ }^\circ\text{C}$.
35. The way of preparing the samples for electrophoretic separation strongly depends on the machinery that is used for this purpose. A polyacrylamide gel matrix or a capillary electrophoresis system is usually used for this purpose and separation takes place according to the specifications of the electrophoretic apparatus that is used.

4.3.4. SSCP (Single Strand Conformation Polymorphism)

The SSCP method relies on the fact that the mobility of single-stranded DNA fragments in a non-denaturing gel is distinctly affected by very small changes in nucleic acid sequence. These small differences become evident because of the relatively unstable nature of single-stranded DNA: If single DNA strands are denatured and subsequently renatured, they undergo a three-dimensional folding, where intra-strand base pairing may occur, resulting in loops and folds that give the single DNA strand a unique conformation. This in turn will noticeably affect the

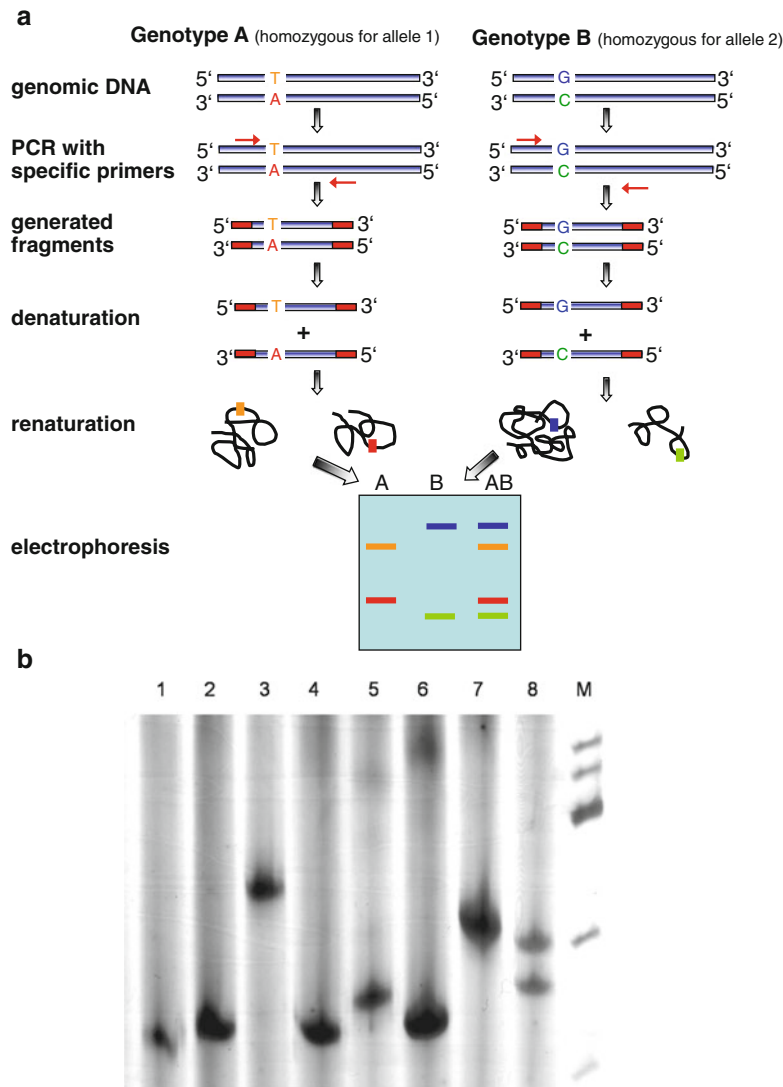


Fig. 4.5 (a) Schematic outline of SSCP analysis. Genomic DNA is amplified with gene-specific primers; amplicons are denatured and renatured prior to electrophoresis. The amplicons with slight differences in sequence take a different conformation when they renature resulting in differences in their migration speed in an electrophoretic gel. (b) Example of a banding pattern obtained after SSCP analysis of eight samples of the entomopathogenic fungus *Beauveria bassiana* based on amplification of part of the beta tubulin gene. M designates a molecular weight marker (picture courtesy of authors).

migration behaviour of the single-stranded DNA fragments since DNA molecules with a different folded structure may migrate faster or slower in a gel matrix depending on their 3D conformation (Fig. 4.5a). SSCP analysis typically begins with the amplification of a fragment of known length in a PCR reaction using specific primers. Some examples of primers commonly used for SSCP analysis are given in Table 4.1. Usually, the length of the

amplification product should be between 100 and 400 bp, since it has been shown that single base pair differences are more accurately resolved for smaller fragments. However, occasionally also fragments of up to 700 or 800 bp have been analysed successfully. After PCR, amplified DNA fragments are subsequently denatured usually by adding a denaturing agent such as formamide, DMSO, or NaOH to an aliquot of the PCR reaction and heating it at 95 °C for 5 min, followed by snap chilling the samples on ice. Electrophoretic separation takes place in a non-denaturing gel, either based on polyacrylamide or a related gel matrix. During electrophoretic separation, a low constant temperature below 10 °C should be maintained, which is usually achieved by connecting the electrophoresis apparatus to a circulating water bath, which automatically makes adjustments in the temperature of the circulating water in order to keep the buffer temperature at a desired level during electrophoresis. SSCP patterns are detected in the gel either by staining the gel in staining agents such as SYBR Gold, ethidium bromide, or silver nitrate or, alternatively, by using an isotopic SSCP protocol. A picture of an SSCP profiling is shown in Fig. 4.5b. Frequently, SSCP products are also subsequently sequenced to locate nucleotide differences between samples in the respective PCR product.

1. Add the following to the tubes on ice (total volume of 50 µl) (see Notes 1, 2, 3):
 - (a) 37.3 µl sterile distilled H₂O
 - (b) 5.0 µl 10× PCR buffer (contains 15 mM MgCl₂)
 - (c) 2.0 µl forward primer (20 pmol/µl)
 - (d) 2.0 µl reverse primer (20 pmol/µl)
 - (e) 1.6 µl 2 mM dNTPs
 - (f) 0.1 µl *Taq* DNA polymerase (5 U/µl)
 - (g) 2.0 µl template DNA (diluted to ca. 20 ng/µl) (see Note 4)
2. Mix gently, centrifuge briefly, and place in thermocycler.
3. PCR program:
 - (a) Initial denaturation at 94 °C for 30 s
 - (b) Amplification at 30 cycles of
 - 94 °C for 30 s
 - 53 °C for 30 s (see Note 6)
 - 72 °C for 1 min
 - (c) Final elongation at 72 °C for 5 min
 - (d) Hold reactions at 4 °C
4. Store reactions at −20 °C.

5. Add 10.0 μl 1 \times formamide dye to the tubes on ice, mix, and centrifuge briefly.
6. Denature samples for 3 min at 95 °C, quickly chill on ice.
7. Load 10–15 μl of the reaction on a polyacrylamide or other non-denaturing gel matrix and separate fragments according to the specifications of the electrophoretic apparatus that is used. Keep temperature constant during the separation at 7–9 °C depending on the requirements of the specific amplification products.

4.3.5. ARDRA (Amplified Ribosomal DNA Restriction Analysis)

The procedure briefly involves the following steps:

1. PCR primers are chosen based on the aim of the experiment. PCR reaction is set up based on the base composition of the primers and the length of the amplicon.
2. Restriction enzyme digestion.
3. Separation of products of the restriction digest on an agarose or polyacrylamide gel.

The detailed procedure includes

1. Add the following to the tubes on ice (total volume of 50 μl) (see Notes 1, 2, 3):
 - (a) 37.3 μl sterile distilled H_2O
 - (b) 5.0 μl 10 \times PCR buffer (contains 15 mM MgCl_2)
 - (c) 2.0 μl forward primer (20 pmol/ μl)
 - (d) 2.0 μl reverse primer (20 pmol/ μl)
 - (e) 5.0 μl 2 mM dNTPs
 - (f) 0.1 μl *Taq* DNA polymerase (5 U/ μl)
 - (g) 2.0 μl template DNA (diluted to ca. 20 ng/ μl) (see Note 4)
2. Mix gently, centrifuge briefly, and place in thermocycler.
3. PCR program:
 - (a) Initial denaturation at 94 °C for 5 min
 - (b) Amplification at 30 cycles of
 - 94 °C for 30s
 - 53 °C for 30 s (see Note 6)
 - 72 °C for 2 min
 - (c) Final elongation at 72 °C for 5 min
 - (d) Hold reactions at 4 °C
4. Store reactions at –20 °C.

5. Run a 5 μ l aliquot of the amplified DNA with 3.0 μ l 6 \times loading dye on a 1 % agarose gel to check amplification.
6. Load 4 μ l of a DNA size marker.
7. Separate fragments at 80 V for ca. 1 h.
8. If amplification is successful, purify the PCR reaction mixture with a commercial PCR purification.
9. Pipette together:
 - (a) ≤ 43 μ l of the PCR product
 - (b) 5.0 μ l 10 \times restriction buffer appropriate for restriction enzyme (chosen)*
 - (c) 2.0 μ l restriction enzyme (20 U/ μ l)
 - (d) Fill up to a total volume of 50 μ l with sterile distilled H₂O
10. Mix gently, centrifuge briefly, and incubate at the temperature and time recommended for the restriction enzyme used.
11. Inactivate restriction enzymes by incubating 15 min at 65 °C, and place on ice (or hold at 4 °C).
12. Separate fragments on a 2 % agarose or a polyacrylamide gel.
13. Document the image for analysis.

*Restriction enzyme with highest number of recognition sequences in the sequence amplified is chosen. When the sequence is known, restriction enzymes can be identified with a software (such as NEBcutter at <http://tools.neb.com/NEBcutter2>).

4.3.6. T-RFLP (Terminal Restriction Fragment Length Polymorphism)

This profiling technique is based on the position of a restriction site closest to a labelled end of an amplified gene. It is a useful modification of ARDRA technique in that only one (terminal) fragment is observed while several fragments are observed in ARDRA technique. The amplification on a DNA sample is performed with one or both the primers having their 5' end labelled with a fluorescent molecule. When both primers are labelled, different coloured fluorescent dyes are required. The fluorescent dyes usually used are 6-FAM, ROX, TAMARA, and HEX. The most commonly used dye is 6-FAM. The PCR products are digested with a restriction enzyme (a 4 bp-cutter) and the fragments are separated using either capillary or polyacrylamide electrophoresis in a DNA sequencer. Only the fluorescently labelled terminal fragments are visible on a sequencer, while the rest of the fragments is not seen (Fig. 4.6). The T-RFLP profile is a graph (electropherogram), with the X-axis representing the size of the fragments and the Y-axis the intensity of the fluorescence of each fragment. What appears on an electrophoresis gel as a band is represented as a peak on the electropherogram. It is assumed

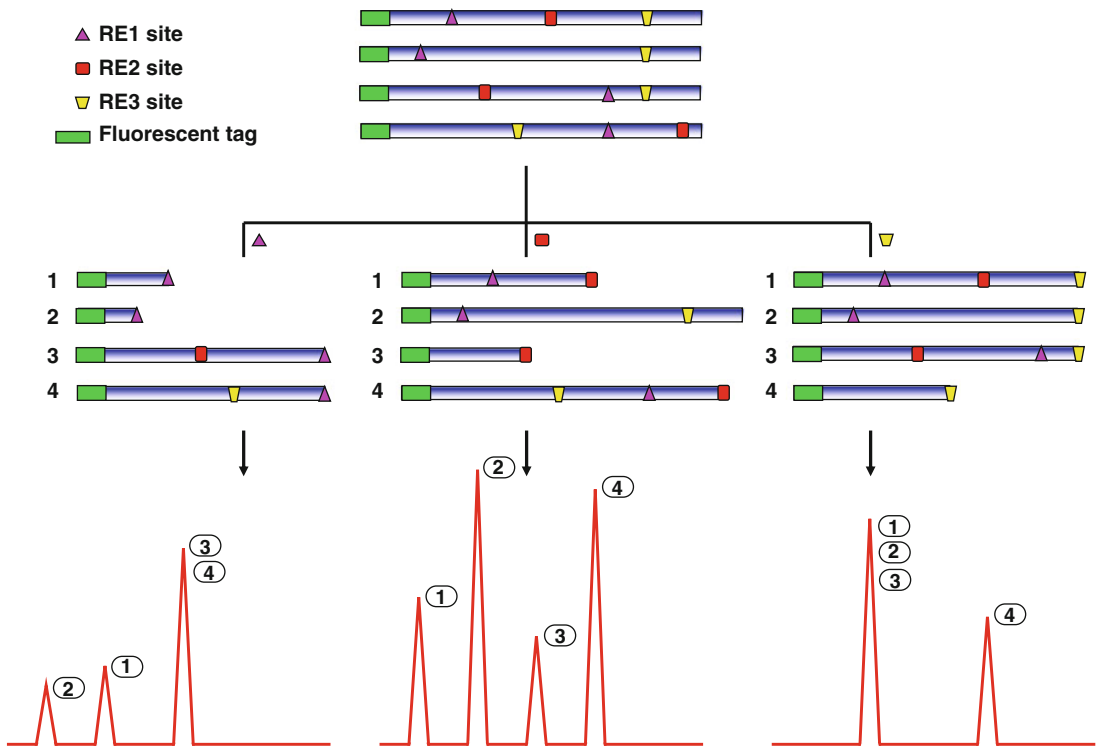


Fig. 4.6 Schematic outline of T-RFLP. Sample DNA is amplified with gene-specific primers, one of the primers being fluorescently labelled. The amplified products are digested with different restriction enzymes and the terminal fragments are detected due to the fluorescent tag in the form of a peak on the electropherogram emerging in an automatic DNA sequencer.

that in a T-RFLP profile, each peak corresponds to one genetic variant in the original sample and its height/area to its relative abundance in that community. But these assumptions are not always correct. Often, several different microbes in the population may have a restriction site for the restriction enzyme used in the experiment at the same position and thus give a single peak on the electropherogram. To overcome this problem and to increase the resolving power of this technique, aliquots of the PCR product can be digested simultaneously with many (often three) enzymes resulting in many T-RFLP profiles for a single sample. Each of these profiles can resolve some variants while missing others. The resolving power of the T-RFLP profiles can be further enhanced by fluorescently labelling the reverse primer also with a different dye. This way, two parallel profiles can be generated per sample each resolving a different number of variants. Usually the PCR products are treated with an exonuclease like mung bean nuclease before proceeding for restriction enzyme digestion to prevent the formation of pseudo-T-RFLPs (arising from random pairing of

ssDNA in the PCR products). Replicate experiments are done to recognise the false background (noise) peaks.

1. Add the following to the tubes on ice (total volume of 50 μ l) (see Notes 1, 2, 3):
 - (a) 5.0 μ l 10 \times PCR buffer (contains 15 mM MgCl₂)
 - (b) 2.0 μ l forward primer (5 pmol/ μ l) (see Note 5)
 - (c) 2.0 μ l forward primer (5 pmol/ μ l)
 - (d) 5.0 μ l 2 mM dNTPs
 - (e) 32.8 μ l sterile distilled H₂O
 - (f) 0.2 μ l *Taq* DNA polymerase (5 U/ μ l)
 - (g) 1.0 μ l template DNA (diluted to ca. 10 ng/ μ l) (see Note 4)
2. Mix gently, centrifuge briefly, and place in thermocycler.
3. PCR program:
 - (a) Initial denaturation at 94 °C for 2 min
 - (b) Amplification at 45 cycles of
 - 94 °C for 1 min
 - 50 °C for 1 min (see Note 6)
 - 72 °C for 2 min
 - (c) Final elongation at 72 °C for 4 min
 - (d) Hold reactions at 4 °C
4. To check for amplification, run 5 μ l of the PCR reaction mix on a 1.5 % agarose gel at 80 V for 1 h. If amplification looks good, proceed to next step.
5. Clean the PCR products with a suitable kit.
6. Concentrate the pooled PCR reactions to a fifth of the original volume using a Speedvac or ethanol precipitation.
7. Pipette together for a 50 μ l volume
 - (a) “x vol” μ l DNA (with 100–150 ng μ g DNA l)
 - (b) 5.0 μ l 10 \times restriction enzyme buffer
 - (c) 2.0 μ l restriction enzyme* (20 U/ μ l)
 - (d) Fill up to a total volume of 50 μ l with sterile distilled H₂O
8. Mix gently, centrifuge briefly, and incubate at the temperature and time recommended for the respective restriction enzyme.
9. Inactivate restriction enzymes by incubating 15 min at 65 °C, and place on ice (or hold at 4 °C).

*Various restriction enzymes can be used in single-enzyme reactions in order to determine which one yields the highest number and most even distribution of terminal restriction fragments. Alternatively if the sequence of the amplified product is

known, software (such as NEBcutter at <http://tools.neb.com/NEBcutter2>) can be used to choose the enzymes.

10. Load an aliquot of the restriction digest adding appropriate volume of loading dye on a polyacrylamide gel. The polyacrylamide gel is cast as per the instructions given for the automatic DNA sequencer being used. In case capillary electrophoresis is used, a desalting step is also done before running the sample. The run is carried out at temperature and time recommended by the manufacturer of the automatic DNA sequencer.
11. The electropherogram is analysed using appropriate software. The analysis can be complemented with a clone library to validate the peaks in the electropherogram and to assess relative abundance of each variant in the library.

4.3.7. DGGE and TGGE (Denaturing/ Temperature Gradient Gel Electrophoresis)

A given sequence of DNA from different individuals within a species may have small differences in sequence. These small (even a single nucleotide) differences can result in a change in the melting point of the sequence. Melting temperature of a DNA molecule is a function of its GC content—the more it is, the higher the denaturing temperature. In DGGE and TGGE, a specific region of DNA is amplified through PCR. The forward primer of the region to be amplified has a 40 nucleotide (GC) long sequence referred to as the GC clamp. When the PCR products from slightly different templates are run on a denaturing gradient gel, the migration of the DNA molecule stops when it is denatured. The GC clamp at one end of the amplified DNA prevents the denatured strands from coming apart. Thus PCR products of the same size but from different individuals within a species or different species (and therefore with differences in sequence) end up at different places on the gel (Fig. 4.7). If species-specific probes are available, individual bands on the gel can be identified by subsequent Southern blotting and hybridisation. Very commonly the conserved regions of the ribosomal genes (16S rRNA, 18S rRNA, or ITS; see Table 4.1) are targeted for PCR. These techniques are not suitable when the sequence is exceptionally GC rich and when the length of it is more than 400 bp.

1. Add the following to the tubes on ice (total volume of 50 μ l) (see Notes 1, 2, 3):
 - (a) 37.3 μ l sterile distilled H₂O
 - (b) 5.0 μ l 10 \times PCR buffer (contains 15 mM MgCl₂)
 - (c) 2.0 μ l forward primer (20 pmol/ μ l)
 - (d) 2.0 μ l reverse primer (20 pmol/ μ l)
 - (e) 5.0 μ l 2 mM dNTPs
 - (f) 0.2 μ l *Taq* DNA polymerase (5 U/ μ l)
 - (g) 2.0 μ l template DNA (diluted to ca. 20 ng/ μ l) (see Note 4)

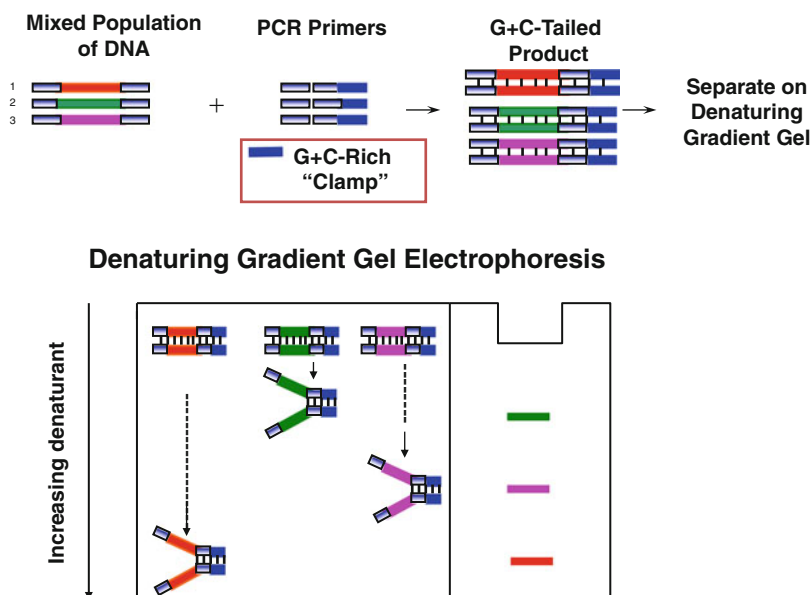


Fig. 4.7 Schematic outline of D/TGGE. Sample DNA is amplified with gene-specific primers; one of the primers has a ca. 40 bp long GC tail. The amplified products are run on a gel with a denaturing gradient (created by the gel constitution or through temperature). The migrating DNA molecules denature when subjected to the denaturing condition in the gel. The site on the gel where they completely denature/melt (except for the GC clamp) is a property of their sequence. The DNA molecules stop migration when they have melted completely. Slight differences in sequence result in differences in the melting point of the DNA molecules and hence the distance they travel on a gel with denaturing gradient.

2. Mix gently, centrifuge briefly, and place in thermocycler
3. PCR program:
 - (a) Initial denaturation at 94 °C for 30 s
 - (b) Amplification at 30 cycles of
 - 94 °C for 30s
 - 53 °C for 30 s (see Note 6)
 - 72 °C for 1 min
 - (c) Final elongation at 72 °C for 5 min
 - (d) Hold reactions at 4 °C
4. Store reactions at –20 °C
5. Check for amplification by running 5 µl on a 1 % agarose gel.
6. For DGGE, prepare the polyacrylamide gel with [8 % (w/v) acrylamide stock solutions (acrylamide/bisacrylamide ratio of 37.5:1) in 1 × TAE buffer (pH 8.0)] with a denaturing gradient of 35–60 % [100 % denaturant contains 7 M urea and 40 % (v/v) formamide]. Maintain the pour rate of the gel at 4 ml/min for an 18 × 16 cm gel. Add 3 ml of stacking polyacrylamide gel with no denaturant after the denaturing gel polymerised for 10 min.

7. For TGGE, prepare the polyacrylamide gel as per the instructions of the apparatus manufacturer. The apparatus has provision to create a temperature gradient with increasing gel from the wells to the bottom.
8. Load 10–15 μl of the reaction on a polyacrylamide gel and separate fragments according to the specifications of the electrophoretic apparatus.
9. Stain the gel for 20 min with ethidium bromide or SYBR Green. Wash the gel twice for 5 min with Milli-Q water and view on a UV transilluminator to capture the picture in digital form.
10. Analyse the gel picture according to the aim of the experiment using an appropriate software. The analysis can be supplemented by further sequencing of the alleles identified in gels.

4.3.8. RFLP **(Restriction Fragment** **Length Polymorphism)**

Due to spontaneous changes in DNA sequence, restriction enzyme sites are either created or lost in different regions of the genome. In a given length of DNA, the difference in the position of the restriction enzymes can be recognised when the restriction digest of DNA is hybridised by a sequence matching that sequence (Fig. 4.8a). An example picture of a RFLP fingerprint is given in Fig. 4.8b.

1. Pipette together in a 20 μl volume
 - (a) “x vol” μl DNA (3–5 μg DNA)
 - (b) 5.0 μl 10 \times restriction enzyme buffer
 - (c) 2.0 μl restriction enzyme (20 U/ μl)
 - (d) Fill up to a total volume of 20 μl with sterile distilled H_2O
2. Mix gently, centrifuge briefly, and incubate at the temperature and time recommended for the enzyme. It would be ideal to leave digestion overnight.
3. Inactivate restriction enzymes by incubating 15 min at 65 $^{\circ}\text{C}$, and place on ice (or hold at 4 $^{\circ}\text{C}$).
4. Cast a 0.6–1 % (depending on the expected size range of the fragments) agarose gel. Use wide thin combs to accommodate a large volume of the sample.
5. Load the restriction digest (maximum volume that can fit in the well) with 3.0 μl 6 \times loading dye.
6. Load 4 μl of DNA size marker (choose a marker suitable for the expected size of the fragments).
7. Run the gel overnight at 70 V (about 16 h) in 1 \times TBE running buffer. Capture the picture of the gel on a UV transilluminator putting a ruler beside the gel in order to estimate the distance run by the size marker.
8. To denature DNA in the gel, place it in 0.25 M HCl for 20 min.

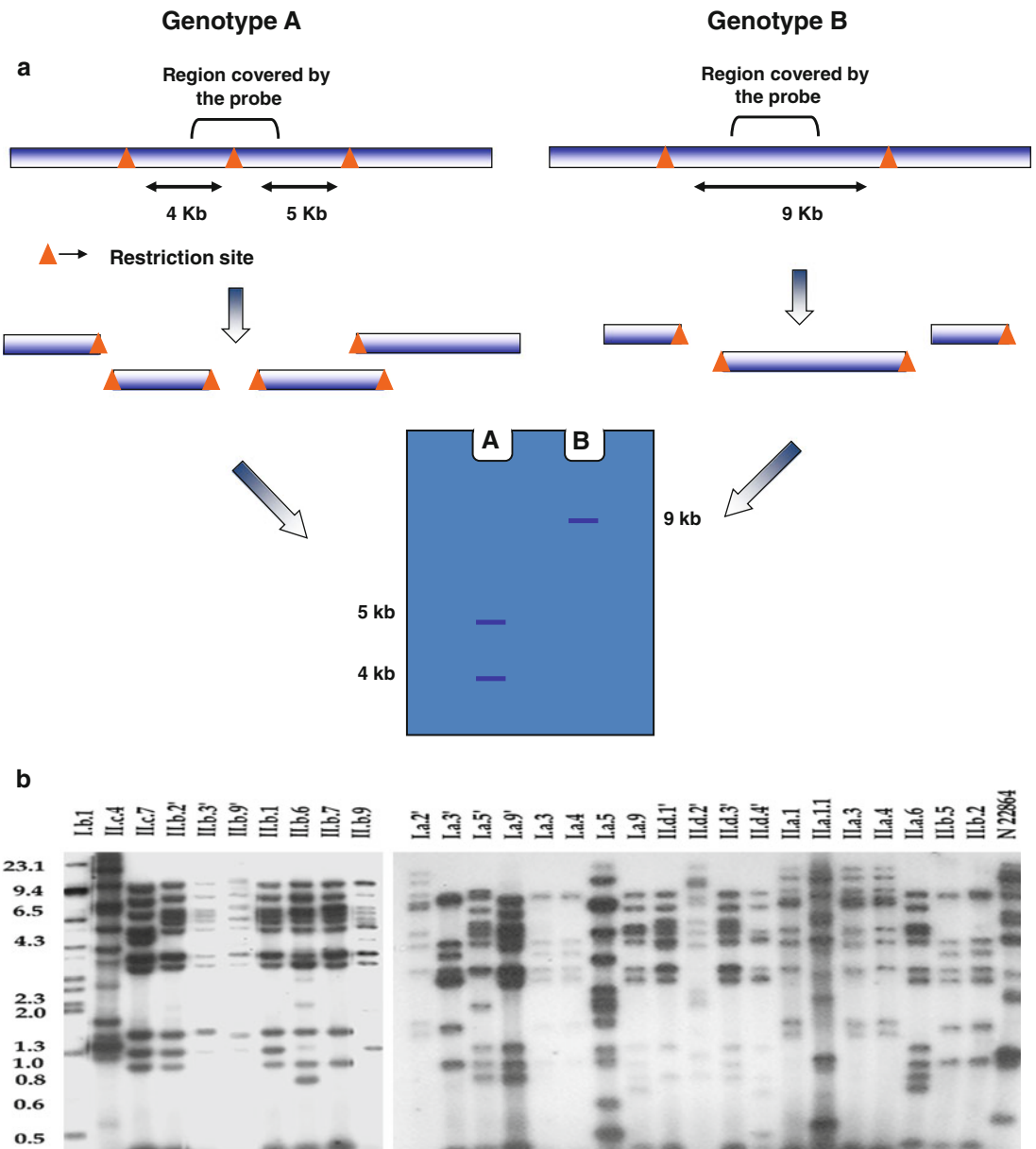


Fig. 4.8 (a) Schematic outline of RFLP. Genomic DNA is digested with a restriction enzyme, restricted products are separated on a gel through electrophoresis, and the separated fragments of DNA on the gel are southern blotted to a membrane and hybridised with a labelled probe sequence. The region spanned by the probe may be in one, two, or three fragments depending on the number of recognition sites of the restriction enzyme used. The fragments that hybridise with the probe light up because of the label and the polymorphism in the locus represented by the probe DNA in a population can be detected. (b) Example of an RFLP used for generating telomere fingerprints. RFLP fingerprints were generated from the *EcoRI*-digested genomic DNA of the entomopathogenic fungi *Nomuraea rileyi* and *Beauveria bassiana* using a p32 end labelled oligonucleotide of telomere repeat sequence (5'-TTTAGG-3')₄ as a probe. Size of molecular weight marker bands is given at the left.

9. Meanwhile set up the vacuum blot apparatus.
 10. Wash the gel with millipore water and place it adjacent to the membrane in blot apparatus. Leave the set-up for 1 h, and check from time to time that there is no leakage. If there is a small leakage, add more NaOH to maintain the gel always submersed. At the end, suck out NaOH. Mark the positions of the wells on the membrane with a pencil.*
 11. Wash the membrane with $2\times$ SSC to clean the agarose.
 12. Expose the membrane to UV light in a UV fixer apparatus (for the time recommended).
 13. Air dry it and wrap with an aluminium foil and store in a $-20\text{ }^{\circ}\text{C}$.
- * If a vacuum blotting apparatus is not available, capillary transfer can be made according to [36].

14. Label the probe DNA (25–50 ng) using a method of choice (end or strand labelling and radio- or nonradioactive label) as per the instructions of the labelling kit.

If the membrane that is used is employed for the first time, an overnight pre-hybridisation is needed, or else 4–5 h is sufficient.

15. Wet the membrane with either distilled water or $2\times$ SSC. Drain off the excess water or SSC. Roll the membrane into the hybridisation tube. Pour 20 ml of the pre-hybridisation solution (see below) along with 600 μg of sonicated and denatured (boiled for 5 min) salmon sperm DNA (10 $\mu\text{g}/\mu\text{l}$) and incubate in the rotating hybridisation oven at $65\text{ }^{\circ}\text{C}$. After pre-hybridisation, start hybridisation by adding to the pre-hybridisation solution the labelled probe (denatured by boiling at $100\text{ }^{\circ}\text{C}$ for 10 min and placing on ice immediately). Incubate overnight.

If radioactive label is used, pour out all the wash solutions in a container meant for disposal of radioactive waste. The stringency (concentration of the wash solution and temperature of incubation of the remaining washings) is decided based on the base composition of the probe. A general protocol is given here.

16. Place the hybridised membrane in a plastic box. Pour 400 ml of $2\times$ SSC into it.
17. Place it on a shaker and shake at medium speed at room temperature for 10 min. Pour off the solution.
18. Add 500 ml of preheated ($65\text{ }^{\circ}\text{C}$) $1\times$ SSC + 0.1 % SDS and incubate on a slow shaking platform for 20 min at $65\text{ }^{\circ}\text{C}$. Pour off the solution and replace with preheated ($65\text{ }^{\circ}\text{C}$) $0.2\times$ SSC + 0.1 % SDS and incubate with shaking for 20 min at $65\text{ }^{\circ}\text{C}$. If radioactive label is used, measure the radioactivity level on the membrane with a Geiger Muller counter. If the count is more than 5 cpm, wash the membrane again at $65\text{ }^{\circ}\text{C}$ with $0.1\times$ SSC + 0.1 % SDS for 20 min.

19. Dry the membrane and expose to an X-ray film. Store the X-ray cassette in a freezer at -80°C .
20. Develop the X-ray film.

4.4 Notes

4.4.1. Some General Considerations for PCR-Based DNA Fingerprinting Methods

1. For setting up PCR reactions, prepare the master mix (minus DNA) of a volume sufficient for all reactions. Add the components (dNTP mixture, primers, PCR buffer, and water) of the master mix to a tube placed on ice. Always add *Taq* DNA polymerase as the last component to the master mix. Mix the master mix components by gentle shaking and spin the contents down through pulse centrifuging. Dispense aliquots of the master mix into PCR tubes/wells (placed on ice) and then add the respective template DNA. Place the tubes or wells in the thermocycler and proceed with the PCR program.
2. Watch out to close tubes or plates properly to prevent evaporation!
3. Always include a negative control (with no DNA) in the PCR reactions.
4. Test reproducibility of PCR reactions by using at least two independently isolated DNA templates from each sample and repeating some of the PCR reactions if possible using different thermocyclers.
5. If fluorescently labelled primers are used (e.g. for AFLP analysis), cover them and subsequent reactions with foil as much as possible to avoid prolonged exposure to light!
6. Most of the annealing temperatures presented in the above protocols are just an example; use an appropriate temperature according to the specifications of the primers used.

References

1. Head IM, Saunders JR, Pickup RW (1998) Microbial evolution, diversity, and ecology: a decade of ribosomal RNA analysis of uncultivated microorganisms. *Microb Ecol* 35:1–21
2. White TJ, Bruns T, Lee S et al (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) *PCR protocols: a guide to methods and applications*. Academic, San Diego, pp 315–322
3. Zinger L, Gury J, Alibeu O et al (2008) CE-SSCP and CE-FLA, simple and high-throughput alternatives for fungal diversity studies. *J Microbiol Methods* 72:42–53
4. Glass NL, Donaldson GC (1995) Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous Ascomycetes. *Appl Environ Microbiol* 61:1323–1330
5. Devi KU, Reineke A, Rao UC et al (2007) AFLP and single-strand conformation polymorphism studies of recombination in the entomopathogenic fungus *Nomuraea rileyi*. *Mycol Res* 111:716–725

6. Weisburg WG, Barns SM, Pelletier DA et al (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 173:697–703
7. Sun L, Qiu F, Zhang X et al (2007) Endophytic bacterial diversity in rice (*Oryza sativa* L.) roots estimated by 16S rDNA sequence analysis. *Microb Ecol* 55:415–424
8. Medlin L, Elwood HJ, Stickel S et al (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* 71:491–499
9. Stoeck T, Epstein S (2003) Novel eukaryotic lineages inferred from small-subunit rRNA analyses of oxygen-depleted marine environments. *Appl Environ Microbiol* 69:2657–2663
10. Borneman J, Hartin RJ (2000) PCR primers that amplify fungal rRNA genes from environmental samples. *Appl Environ Microbiol* 66:4356–4360
11. Edel-Hermann W, Dreumont C, Perez-Piqueres A et al (2004) Terminal restriction fragment length polymorphism analysis of ribosomal RNA genes to assess changes in fungal community structure in soils. *FEMS Microbiol Ecol* 47:397–404
12. Lane DJ (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds) *Nucleic acid techniques in bacterial systematics*. Wiley, West Sussex, pp 115–175
13. Dees PM, Ghiorse WC (2001) Microbial diversity in hot synthetic compost as revealed by PCR-amplified rRNA sequences from cultivated isolates and extracted DNA. *FEMS Microbiol Ecol* 35:207–216
14. Schütte UME, Abdo Z, Bent SJ et al (2008) Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Appl Microbiol Biotechnol* 80:365–380
15. Gothwal RK, Nigam VK, Mohan MK et al (2007) Extraction of bulk DNA from Thar Desert soils for optimization of PCR-DGGE based microbial community analysis. *Electron J Biotechnol* 10:400–408
16. Muyzer G, de Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 59:695–700
17. Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* 18:7213–7218
18. Williams JGK, Kubelik AR, Livak KJ et al (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18:6531–6535
19. Meunier JR, Grimont PAD (1993) Factors affecting reproducibility of random amplified polymorphic DNA fingerprinting. *Res Microbiol* 144:373–379
20. Muralidharan K, Wakeland EK (1993) Concentration of primer and template qualitatively affects products in random-amplified polymorphic DNA PCR. *Biotechniques* 14:362–364
21. Schierwater B, Ender A (1993) Different thermostable DNA polymerases may amplify different RAPD products. *Nucleic Acids Res* 21:4647–4648
22. Vos P, Hogers R, Bleeker M et al (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23:4407–4414
23. Meudt HM, Clark AC (2006) Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends Plant Sci* 12:106–117
24. Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. *Trends Ecol Evol* 14:389–394
25. Savelkoul PHM, Aarts HJM, de Haas J et al (1999) Amplified-fragment length polymorphism analysis: the state of an art. *J Clin Microbiol* 37:3083–3091
26. Goldstein DB, Schlötterer C (1999) Microsatellites: evolution and applications. Oxford University Press, Oxford
27. MacGregor BJ, Amann R (2006) Single-stranded conformational polymorphism for separation of mixed rRNAs (rRNA-SSCP), a new method for profiling microbial communities. *Syst Appl Microbiol* 29:661–670
28. Oto M, Suda W, Shinoyama H (2006) 16S rRNA gene-based analysis of microbial community by whole-genome amplification and minigel-single-strand conformation polymorphism technique. *J Biosci Bioeng* 102:482–484
29. Sunnucks P, Wilson ACC, Beheregaray LB et al (2000) SSCP is not so difficult: the application and utility of single-strand conformation polymorphism in evolutionary biology and molecular ecology. *Mol Ecol* 9:1699–1710
30. Zinger L, Gury J, Giraud F et al (2007) Improvements of polymerase chain reaction and capillary electrophoresis single-strand conformation polymorphism methods in microbial ecology: toward a high-throughput method for microbial diversity studies in soil. *Microb Ecol* 54:203–216
31. Gich FB, Amer E, Figueras JB et al (2000) Assessment of microbial community structure changes by amplified ribosomal DNA restriction analysis ARDRA. *Int Microbiol* 3:103–106
32. Wu XY, Walker MJ, Hornitzky M et al (2006) Development of a group-specific PCR combined with ARDRA for the identification of

- Bacillus* species of environmental significance. J Microbiol Methods 64:107–119
33. Liu WT, Marsh TL, Cheng H et al (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl Environ Microbiol 63:4516–4522
 34. de Souza FA, Kowalchuk GA, Leeﬂang P et al (2004) PCR-denaturing gradient gel electrophoresis proﬁling of inter- and intraspecies 18S rRNA gene sequence heterogeneity is an accurate and sensitive method to assess species diversity of arbuscular mycorrhizal fungi of the genus *Gigaspora*. Appl Environ Microbiol 70:1413–1424
 35. Etscheid M, Riesner D (1998) TGGE and DGGE. In: Karp A, Isaac PG, Ingram DS (eds) Molecular tools for screening biodiversity: plants and animals. Chapman and Hall, London, pp 135–151
 36. Sambrook J, Russel DW (2001) Molecular cloning: a laboratory manual, vol 3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
 37. Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11:1–16
 38. Hamilton MB, Pincus EL, Di Fiore A et al (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. Biotechniques 27:500–507
 39. Nunome T, Negoro S, Miyatake K et al (2006) Protocol for the construction of microsatellite enriched genomic library. Plant Mol Biol Rep 24:305–312

Agarose Gel Electrophoresis and Polyacrylamide Gel Electrophoresis: Methods and Principles

Sukumar Mesapogu, Chandra Mouleswararao Jillepalli,
and Dilip K. Arora

Abstract

Electrophoresis is a technique used to separate and sometimes purify macromolecules—especially proteins and nucleic acids—that differ in size, charge, or conformation. When charged molecules are placed in an electric field, they migrate toward either the positive or negative pole according to their charge. In contrast to proteins, which can have either a net positive or net negative charge, nucleic acids have a consistent negative charge imparted by their phosphate backbone and migrate toward the anode. Proteins and nucleic acids are electrophoresed within a matrix or gel. The gel is immersed within an electrophoresis buffer that provides ions to carry a current and some type of buffer to maintain the pH at a relatively constant value. Agarose is typically used at concentrations of 0.5–2 %. Agarose gels have a large range of separation, but relatively low resolving power. By varying the concentration of agarose, fragments of DNA from about 200 to 50,000 bp can be separated using standard electrophoretic techniques. SDS PAGE uses an anionic detergent (SDS) to denature proteins and the protein molecules become linearized. One SDS molecule binds to two amino acids. Due to this, the charge to mass ratio of all the denatured proteins in the mixture becomes constant. These protein molecules move in the gel (toward the anode) on the basis of their molecular weights only and are separated. The polyacrylamide chains are cross linked by *N,N*-methylene bisacrylamide comonomers. Polymerization is initiated by ammonium persulfate (radical source) and catalyzed by TEMED.

5.1 Introduction

Electrophoresis is a procedure which enables the sorting of molecules based on size and charge. Using an electric field, molecules (such as DNA) can be made to move through a gel made of agar or polyacrylamide. Nucleic acid molecules are separated by applying an electric field to move the negatively charged molecules through an agarose matrix. Shorter molecules move faster and migrate farther than longer ones because shorter molecules migrate more easily through the pores of the gel. This phenomenon is called sieving. Proteins are separated by charge in agarose because the pores of the gel are too large to sieve proteins. Gel electrophoresis can also be used for separation of nanoparticles [1].

The term “gel” in this instance refers to the matrix used to contain and then separate the target molecules. In most cases, the gel is a cross-linked polymer whose composition and porosity is chosen based on the specific weight and composition of the target to be analyzed. When separating proteins or small nucleic acids (DNA, RNA, or oligonucleotides) the gel is usually composed of different concentrations of acrylamide and a cross-linker, producing different sized mesh networks of polyacrylamide. When separating larger nucleic acids (greater than a few hundred bases), the preferred matrix is purified agarose. In both cases, the gel forms a solid, yet porous matrix. Acrylamide, in contrast to polyacrylamide, is a neurotoxin and must be handled using appropriate safety precautions to avoid poisoning. Agarose is composed of long unbranched chains of uncharged carbohydrate without cross-links resulting in a gel with large pores allowing for the separation of macromolecules and macromolecular complexes. DNA Gel electrophoresis is usually performed for analytical purposes, often after amplification of DNA via PCR, but may be used as a preparative technique prior to use of other methods such as mass spectrometry, RFLP, PCR, cloning, DNA sequencing, or Southern blotting for further characterization [2].

Electrophoresis apparatus is arguably one of the most vital pieces of equipment in the laboratory. It consists of four main parts: a power supply (capable of at least 100 V and currents of up to 100 mA), an electrophoresis tank, a casting plate, and a well-forming comb. Apparatus is available from many commercial suppliers, but tends to be fairly expensive. The essence of electrophoresis is that when DNA molecules within an agarose gel matrix are subjected to a steady electric field, they first orient in an end-on position and then migrate through the gel at rates that are inversely proportional to the log of the number of base pairs. This is because larger molecules migrate more slowly than smaller molecules because of their higher frictional drag and greater difficulty in “worming” through the pores of the gel. This relationship only applies to linear molecules. Circular molecules, such as plasmids, migrate much more quickly than their molecular weight would imply because of their smaller apparent size with respect to the gel matrix [3].

There are limits to electrophoretic techniques. Since passing current through a gel causes heating, gels may melt during electrophoresis. Electrophoresis is performed in buffer solutions to reduce pH changes due to the electric field, which is important because the charge of DNA and RNA depends on pH, but running for too long can exhaust the buffering capacity of the solution. Further, different preparations of genetic material may not migrate consistently with each other, for morphological or other reasons.

Agarose gels are easily cast and handled compared to other matrices, because the gel setting is a physical rather than chemical change. Samples are also easily recovered. After the experiment is finished, the resulting gel can be stored in a plastic bag in a refrigerator. Agarose gel electrophoresis can be used for the separation of DNA fragments ranging from 50 bp to several megabases (millions of bases) using specialized apparatus. The distance between DNA bands of a given length is determined by the percent agarose in the gel. The migration rate also depends on other factors, such as the composition and ionic strength of the electrophoresis buffer as well as the percentage of agarose in the gel. The gel percentage presents the best way to control the resolution of agarose gel electrophoresis (see Table 5.2). The disadvantage of higher concentrations is the long run times (sometimes days). Instead high percentage agarose gels should be run with a pulsed field electrophoresis (PFE) or field inversion electrophoresis. Most agarose gels are made with between 0.7 % (good separation or resolution of large 5–10 kb DNA fragments) and 2 % (good resolution for small 0.2–1 kb fragments) agarose dissolved in electrophoresis buffer. Up to 3 % can be used for separating very tiny fragments but a vertical polyacrylamide gel is more appropriate in this case. Low percentage gels are very weak and may break when you try to lift them. High percentage gels are often brittle and do not set evenly. 1 % gels are common for many applications [4]. Agarose gels do not have a uniform pore size, but are optimal for electrophoresis of proteins that are larger than 200 kDa [5].

A very common method for separating proteins by electrophoresis uses a discontinuous polyacrylamide gel as a support medium and sodium dodecyl sulfate (SDS) to denature the proteins. The method is called sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). SDS (also called lauryl sulfate) is an anionic detergent; the negative charges on SDS destroy most of the complex structure of proteins and are strongly attracted toward an anode (positively charged electrode) in an electric field. Polyacrylamide gels restrain larger molecules from migrating as fast as smaller molecules. Because the charge-to-mass ratio is nearly the same among SDS-denatured polypeptides, the final separation of proteins is dependent almost entirely on the differences in relative molecular mass of polypeptides. Protein separation by SDS-PAGE can be used to estimate relative molecular mass, to determine the relative abundance of major proteins in a sample, and to determine the distribution of proteins among fractions. The purity of protein samples can be assessed and the progress of a fractionation or purification procedure can be followed. Different staining methods can be used to detect rare proteins and to learn something about their biochemical properties.

Many systems for protein electrophoresis have been developed, and apparatus used for SDS-PAGE varies widely. The methodology used on these pages employs the Laemmli method. SDS-PAGE can be conducted on precast gels, saving the trouble and hazard of working with acrylamide. Regardless of the system, preparation requires casting two different layers of acrylamide between glass plates. The lower layer (separating or resolving, gel) is responsible for actually separating polypeptides by size. The upper layer (stacking gel) includes the sample wells. It is designed to sweep up proteins in a sample between two moving boundaries so that they are compressed (stacked) into micrometer thin layers when they reach the separating gel.

Polyacrylamide gel electrophoresis (PAGE) is used for separating proteins ranging in size from 5 to 2,000 kDa due to the uniform pore size provided by the polyacrylamide gel. Pore size is controlled by controlling the concentrations of acrylamide and bisacrylamide powder used in creating a gel. Care must be used when creating this type of gel, as acrylamide is a potent neurotoxin in its liquid and powdered form. Traditional DNA sequencing techniques such as Maxam-Gilbert or Sanger methods used polyacrylamide gels to separate DNA fragments differing by a single base-pair in length so the sequence could be read. Most modern DNA separation methods now use agarose gels, except for particularly small DNA fragments. It is currently most often used in the field of immunology and protein analysis, often used to separate different proteins or isoforms of the same protein into separate bands. These can be transferred onto a nitrocellulose or PVDF membrane to be probed with antibodies and corresponding markers, such as in a western blot. Typically resolving gels are made in 6 %, 8 %, 10 %, 12 %, or 15 %. Stacking gel (5 %) is poured on top of the resolving gel and a gel comb (which forms the wells and defines the lanes where proteins, sample buffer, and ladders will be placed) is inserted. The percentage chosen depends on the size of the protein that one wishes to identify or probe in the sample. The smaller the known weight, the higher the percentage that should be used. Changes on the buffer system of the gel can help to further resolve proteins of very small sizes [6].

There are a number of buffers used for electrophoresis. The most common being for nucleic acids Tris/Acetate/EDTA (TAE) and Tris/Borate/EDTA (TBE). Many other buffers have been proposed, e.g., lithium borate, which is almost never used, based on Pubmed citations (LB), iso electric histidine, pK matched goods buffers, etc.; in most cases the purported rationale is lower current (less heat) and or matched ion mobilities, which leads to longer buffer life. Borate is problematic; borate can polymerize and/or interact with cis diols such as those found in RNA.

TAE has the lowest buffering capacity but provides the best resolution for larger DNA. This means a lower voltage and more time, but a better product. LB is relatively new and is ineffective in resolving fragments larger than 5 kbp; however, with its low conductivity, a much higher voltage could be used (up to 35 V/cm), which means a shorter analysis time for routine electrophoresis. As low as one base pair size difference could be resolved in 3 % agarose gel with an extremely low conductivity medium (1 mM Lithium borate) [7].

After the electrophoresis, the molecules in the gel can be stained to make them visible. DNA may be visualized using ethidium bromide which, when intercalated into DNA, fluoresce under ultraviolet light, while protein may be visualized using silver stain [8] or Coomassie Brilliant Blue dye. Other methods may also be used to visualize the separation of the mixture's components on the gel. If the molecules to be separated contain radioactivity, for example in DNA sequencing gel, an autoradiogram can be recorded of the gel. Photographs can be taken of gels, often using Gel Doc. The most common dye used to make DNA bands visible for agarose gel electrophoresis is ethidium bromide, usually abbreviated as EtBr. It fluoresces under UV light when intercalated into the major groove of DNA (or RNA). By running DNA through an EtBr-treated gel and visualizing it with UV light, any band containing more than ~20 ng DNA becomes distinctly visible. EtBr is a known mutagen, and safer alternatives are available, such as GelRed, which binds to the minor groove. SYBR Green I is another dsDNA stain, produced by Invitrogen. It is more expensive, but 25 times more sensitive, and possibly safer than EtBr, though there is no data addressing its mutagenicity or toxicity in humans. Since EtBr stained DNA is not visible in natural light, scientists mix DNA with negatively charged loading buffers before adding the mixture to the gel. Loading buffers are useful because they are visible in natural light (as opposed to UV light for EtBr-stained DNA), and they cosediment with DNA (meaning they move at the same speed as DNA of a certain length). Xylene cyanol and Bromophenol blue are common dyes found in loading buffers; they run about the same speed as DNA fragments that are 5,000 bp and 300 bp in length respectively, but the precise position varies with percentage of the gel. Other less frequently used progress markers are Cresol Red and Orange G which run at about 125 bp and 50 bp, respectively. After electrophoresis, the gel is illuminated with an ultraviolet lamp. The ethidium bromide fluoresces reddish-orange in the presence of DNA, since it has intercalated with the DNA. The gel can then be photographed usually with a digital or polaroid camera. Although the stained nucleic acid fluoresces reddish-orange, images are usually shown in black and white. Even short exposure of nucleic acids

to UV light causes significant damage to the sample. UV damage to the sample will reduce the efficiency of subsequent manipulation of the sample, such as ligation and cloning. If the DNA is to be used after separation on the agarose gel, it is best to avoid exposure to UV light by using a blue light excitation source such as the XcitaBlue UV to blue light conversion screen from Bio-Rad or Dark Reader from Clare Chemicals. A blue excitable stain is required, such as one of the SYBR Green or GelGreen stains. Blue light is also better for visualization since it is safer than UV (eye protection is not such a critical requirement) and passes through transparent plastic and glass. This means that the staining will be brighter even if the excitation light goes through glass or plastic gel platforms.

In the case of DNA, polyacrylamide is used for separating fragments of less than about 500 bp. However, under appropriate conditions, fragments of DNA differing in length by a single base pair are easily resolved. In contrast to agarose, polyacrylamide gels are used extensively for separating and characterizing mixtures of proteins.

5.2 Materials

5.2.1. Agarose Gel Electrophoresis

1. Molecular-biology grade agarose (high melting point, see Table 5.1).
2. Running buffer at 1× and 10× concentrations (see Table 5.2 for choice).
3. Sterile distilled water.
4. A heating plate or microwave oven.
5. Suitable gel apparatus and power pack (see Fig. 5.1).
6. Ethidium bromide: dissolve in water at 10 mg/ml (Ethidium bromide is both carcinogenic and mutagenic and therefore must be handled with extreme caution).
7. An ultraviolet (UV) light transilluminator (long wave, 365 nm will not damage the DNA very fast).
8. 5× loading buffer (see Note 2): Many variations exist, but this one is fairly standard: 50 % (v/v) glycerol, 50 mM EDTA, pH 8.0, 0.125 % (w/v) bromophenol blue, 0.125 % (w/v) xylene cyanol.
9. A size marker: a predigested DNA sample for which the product band sizes are known. Many such markers are commercially available.

Table 5.1
Resolution of agarose gels

Agarose %	Mol wt range (kb)	Comments
0.2	5–40	Gel very weak; separation in 20–40 kb range improved by increase in ionic strength of running buffer (i.e., Loenings E); only use high-melting point agarose
0.4	5–30	With care can use low-melting point agarose
0.6	3–10	Essentially as above, but with greater mechanical strength
0.8	1–7	General-purpose gel separation not greatly affected by choice of running buffer, bromophenol blue runs at about 1 kb
1	0.5–5	As for 0.8 %
1.5	0.3–3	As for 0.8 %, bromophenol blue runs at about 500 bp
2.0	0.2–1.5	Do not allow to cool to 50 °C before pouring
3.0	0.1–1	Can separate small fragments differ mg from each other by a small amount; must be poured rapidly onto a prewarmed glass plate

Table 5.2
Commonly used agarose gel electrophoresis running buffers

Buffer	Description	Solution
Loenings E	High ionic strength, and not recommended for preparative gels	For 5 L of 10X: 218 g of Tns base, 234 g of $\text{NaH}_2\text{PO}_4 \cdot 2 \text{H}_2\text{O}$, and 18.6 g of Na, EDTA-2 H_2O
Glycine	Low ionic strength, very good for preparative gels, but can also be used for analytical gels	For 2 L 10X: 300 g of glycine, 300 ml of 1 M NaOH (or 12 g pellets), and 80 ml of 0.5 M EDTA, pH 8.0
Tris-borate EDTA (TBE)	Low ionic strength can be used for both preparative and analytical gels	For 5 L of 10X: 545 g of Tns, 278 g boric acid, and 46.5 g of EDTA
Tris-acetate (TAE)	Good for analytical gels and preparative gels when the DNA is to be purified by glass beads	For 1 L of 50X: \times 242 g of Tns base, 57.1 ml of glacial acetic acid, and 100 ml of 0.5 M EDTA, pH 8.0

5.2.2. SDS-PAGE (Poly Acrylamide Gel Electrophoresis)

1. Stacking Gel Solution: 10 ml of total volume is good for 2 mini gels, so measure out other components and make up to 10 ml final volume with distilled water. Final concentration of acrylamide is 4.44 %.

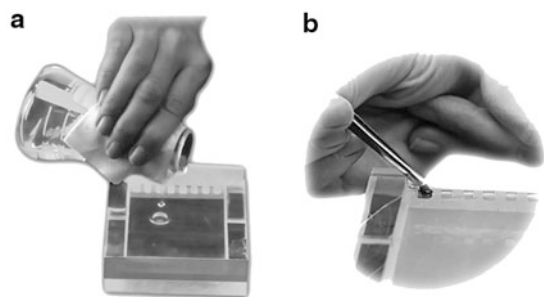


Fig. 5.1 (a) Melting of agarose and dispensing into the casting tray. (b) Loading of the samples

22.2 % Acrylamide/bisacrylamide	2 ml
Distilled water	6.6 ml
1 M Tris/HCl pH 6.8	1.25 ml
10 % SDS	100 l
10 % Ammonium persulfate	50 µl
TEMED	5 µl

2. Solutions needed

22.2 % Acrylamide/Bisacrylamide mix: 22.2 g acrylamide, 0.6 g bis-acrylamide (37:1 cross-linker ratio) to 100 ml water, filtered. (*Acrylamide is a potent neurotoxin and should be handled with care!* Wear disposable gloves when handling solutions of acrylamide and a mask when weighing out powder. Polyacrylamide is considered to be nontoxic, but polyacrylamide gels should also be handled with gloves due to the possible presence of free acrylamide Table 5.3).

44.4 % Acrylamide/Bisacrylamide mix: 44.4 g acrylamide, 1.2 g bis-acrylamide (37:1 cross-linker ratio) to 100 ml water, filtered Table 5.4.

Reservoir/running buffer: 57.6 g Glycine, 12 g Tris base, 4 g SDS, water to 4 l.

Stain solution: 2.5 g Coomassie Brilliant Blue R-250, 450 ml methanol, 100 ml glacial acetic acid, water to 1 liter.

Destain solution: 300 ml methanol, 400 ml acetic acid, water to 4 l.

Sample buffer 5×: make up 100 ml and store away 5–10 ml aliquots.

Table 5.3**Proportions of chemical solution in order to prepare 5–12 % Acrylamide PAGE gels**

Chemicals	12 %	10 %	8 %	7.5 %	6 %	5 %
22.2 % Acrylamide/0.6 % Bis	10.81 ml	9.01 ml	7.21 ml	6.76 ml	5.41 ml	4.5 ml
1 M Tris/HCl pH 8.8	7.5 ml	7.5 ml	7.5 ml	7.5 ml	7.5 ml	7.5 ml
Distilled water	1.38 ml	3.18 ml	4.99 ml	5.43 ml	6.78 ml	7.69 ml
10 % SDS	200 µl	200 µl	200 µl	200 µl	200 µl	200 µl
10 % Ammonium persulfate	100 µl	100 µl	100 µl	100 µl	100 µl	100 µl
TEMED	10 µl	10 µl	10 µl	10 µl	10 µl	10 µl

Table 5.4**Proportions of chemical solution in order to prepare 13.5–27 % Acrylamide PAGE gels**

	27 %	24 %	20 %	17.5 %	15 %	13.5 %
44.4 % Acrylamide/1.2 % Bis	12.16 ml	10.81 ml	9.01 ml	7.88 ml	6.76 ml	6.08 ml
1 M Tris/HCl pH 8.8	7.5 ml	7.5 ml	7.5 ml	7.5 ml	7.5 ml	7.5 ml
Distilled water	0.03 ml	1.38 ml	3.18 ml	4.31 ml	5.43 ml	6.11 ml
10 % SDS	200 µl	200 µl	200 µl	200 µl	200 µl	200 µl
10 % Ammonium persulfate	100 µl	100 µl	100 µl	100 µl	100 µl	100 µl
TEMED	10 µl	10 µl	10 µl	10 µl	10 µl	10 µl

1 M Tris/HCl pH 6.8	31.25 ml
SDS powder	10 g
Glycerol	25 ml
Bromophenol blue (2 % in ethanol)	750 µl
2-Mercaptoethanol	5 µl
Water	100 ml

5.2.3. Silver Staining SDS PAGE Gels

1. Silver nitrate
2. 1 % Citric acid: 100 ml of distilled water + 1 g of citric acid
3. 30 % NaOH (7.5 M): 100 ml of DI water + 30 g of NaOH
4. 14.8 M Ammonium hydroxide
5. 38 % Formaldehyde

6. Ultrapure water, use this for all steps and reagents
7. 50 % Aqueous glutaraldehyde (optional)
8. Glass tray or Novex Stain Ease Gel tray. If using glass, make sure to clean well with soap and DI water

5.3 Method

5.3.1. Agarose Gel Electrophoresis

5.3.1.1. Melting of Agar

1. An appropriate amount of powdered agarose (Table 5.1) is weighed carefully into a conical flask.
2. One-tenth of the final volume of 10× concentrated running buffer is added (Table 5.2), followed by distilled water to the final volume (i.e., 1 ml of 10× buffer in 9 ml of distilled water to make 1× of 10 ml).
3. Cover the container with plastic wrap. Pierce a small hole in the plastic for ventilation.
4. Heat the solution in the microwave oven on high power until it comes to a boil. Watch the solution closely; agarose foams up and boils over easily.
5. Remove the container (protect your hand with a pot holder or folded paper towel) and gently swirl it to resuspend any settled agar.
6. Continue this process until the agar dissolves completely.
7. Cool the agar until you can comfortably touch the flask and add ethidium bromide solution to give a final concentration of 5 pg/ml.
8. The gel mixture is ready to be poured into the gel apparatus (Fig. 5.1a).

5.3.1.2. Pouring the Gel

1. Place tape across the ends of the gel form and place the comb in the form.
2. Pour cooled agar into the form. The agar should come at least half way up the comb teeth.
3. Immediately rinse and fill the agar flask with hot water to dissolve any remaining agar.
4. When the agar has solidified, carefully remove the comb.
5. Remove the tape from the ends of the gel form.

5.3.1.3. Loading the Samples

1. Make a written record of which sample you will load in each well of the gel. You may find it helpful to load samples in every other well.

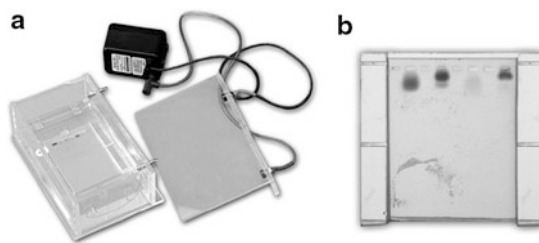


Fig. 5.2 (a) Setting up the gel for electrophoresis. (b) Running and analysis of the samples under UV transilluminator

2. Place the gel form on a black or dark surface to help you see the wells in the agar.
3. Fold filter paper circle in half and hold sideways using tweezers.
4. Dip filter paper into full-strength food color to saturate.
5. Gently ease the filter paper into the well.
6. Be careful not to puncture the bottoms of the wells as you load each sample (Fig. 5.1b).
7. Repeat for remaining colors.

5.3.1.4. Setting up the Gel

1. Place the gel in the electrophoresis chamber.
2. Make sure that the wells are closest to the negative (black) electrode.
3. To prepare the buffer, add 100 ml of 10 \times to 900 ml deionized water or distilled water to make 1 l of 1 \times running buffer (the water source that works for you may depend on your local water quality) and swirl to dissolve.
4. Fill with 1 \times running buffer to just cover the wells.
5. Fill each half of the chamber, adding solution until it is close to the top of the gel. Gently flood the gel from the end opposite the wells to minimize sample diffusion.
6. Place the lid on the chamber and connect the electrode leads to the power supply.
7. Connect the black lead to the negative terminal and the red lead to the positive terminal (Fig. 5.2a).

5.3.1.5. Running and Analyzing the Gel

1. Turn on the power supply and adjust the voltage to 50–100 V.
2. The gel is usually run between 1 and 3 h, depending on the percentage of the gel and length.
3. Once the dyes have moved through the gel, turn off the power supply, disconnect the electrode leads, and remove the chamber lid.

4. Remove the gel from the electrophoresis chamber and analyze your results. Did some colors move further than others? Did some colors separate into two?
5. After electrophoresis, the gel is removed from the apparatus, and the products of the digestion can be viewed on a UV transilluminator (Fig. 5.2b).

5.3.2. SDS-PAGE (Polyacrylamide Gel Electrophoresis)

5.3.2.1. Preparing SDS Gels

A gel of given acrylamide concentration separates proteins effectively within a characteristic range. Very large polypeptides cannot penetrate far into a gel and thus their corresponding bands may be too compressed for resolution. Polypeptides below a particular size are not restricted at all by the gel, and regardless of mass they all move at the same pace along with the tracking dye. Gel concentration (%T) should be selected so that the proteins of interest are resolved.

A typical gel of 7 % acrylamide composition nicely separates polypeptides with molecular mass between 45 and 200 kDa. Polypeptides below the cutoff of around 45 kDa do not resolve. A denser gel, say 14%T, usually resolves all of the smallest polypeptides in a mix. Such a gel would be needed to resolve hemoglobin, for example. It would be useless for resolving bands much above 60 kDa, though. To analyze the entire profile of a fraction that contains heavy and light polypeptides, one should usually run two gels.

5.3.2.2. Cassettes

There are many systems for setting up gel cassettes, A simple ‘mini-slab’ gel system can be put together for a surprisingly little amount of money and does the job quite well. We use casting stands to prepare the mini-slab gels. Two clean plates with two teflon spacers make a single cassette. Stack the cassettes upright in the stand with the bottoms of the cassettes tight to the bottom of the stand, using modeling clay to seal a thick acrylic cover in place against the last cassette to make a water-tight chamber. Using a well-former (comb) as a template, mark a fill line about a centimeter of the first (separating) gel solution (Fig. 5.3).

5.3.2.3. Separating Gel Preparation

The total volume between the plates of our gel cassettes is 10 ml, so if we prepare 10 ml separating gel mix per cassette we have more than enough. From 30 % acrylamide stock (see notes below), we prepare gels of composition 7–15 % acrylamide, depending on the range of proteins that we wish to separate. Our separating gel buffer stock (4× concentrated) consists of 0.4 % SDS, 1.5 M Tris-Cl, pH 8.8. Per cassette, mix 2.5 ml buffer stock and sufficient acrylamide stock so that the mix is brought to final volume with distilled water.

Acrylamide polymerizes spontaneously in the absence of oxygen, so the polymerization process involves complete removal

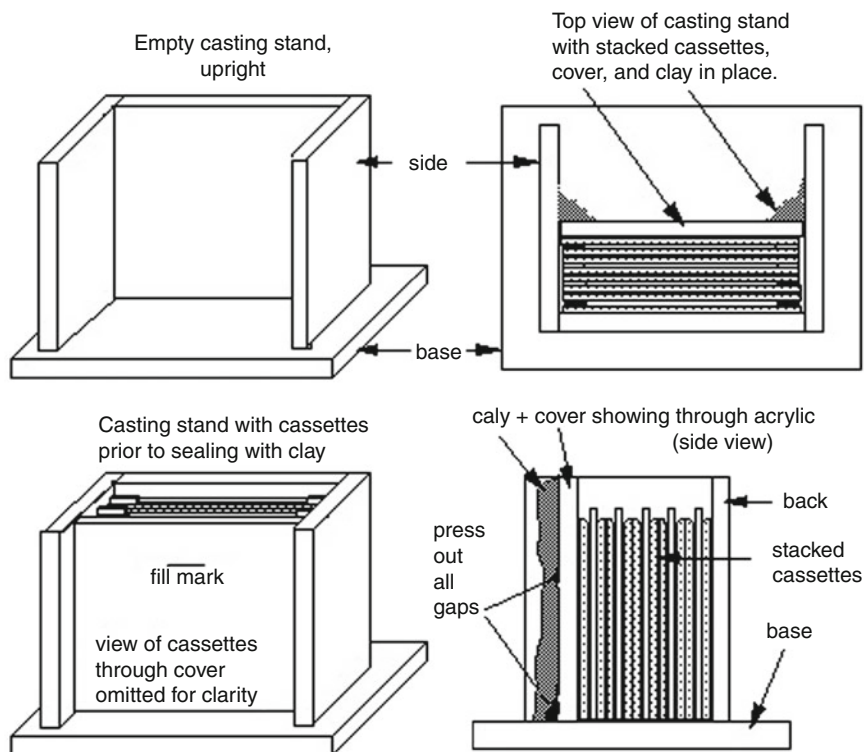


Fig. 5.3 A single gel cassette and assembly

of oxygen from the solution. Polymerization is more uniform if the mix is degassed to remove much of the dissolved oxygen, by placing it under a vacuum for 5 min or so before polymerization. Polymerization by adding freshly prepared 10% ammonium persulfate (AP) to the mix followed by N,N,N',N' -tetramethylethylenediamine (TEMED). The amounts of each depend on the quality of acrylamide used and should be determined in advance by trial and error. Usually 100 μ l AP and 10 μ l TEMED per 10 ml gel mix, once the catalysts are added, polymerization may occur quickly; thus it is necessary to have the casting stand completely ready and to have the overlay solution ready to go. After swirling, mix the solution into the space occupied by the cassettes. The cassettes will self-level eventually, but leveling can be hurried along by adding solution to selected cassettes with a pasteur pipet.

Immediately after pouring the gel mix, it must be overlaid with water-saturated butanol to an additional height of 0.5 cm or so. The purpose of butanol is to produce a smooth, completely level surface on top of the separating gel, so that bands are straight and uniform. Butanol holds very little water in solution, forming a neat layer on top. Water would make an effective overlay but would mix with the acrylamide solution diluting it. Polymerization can be confirmed by pulling some of the remaining gel mix

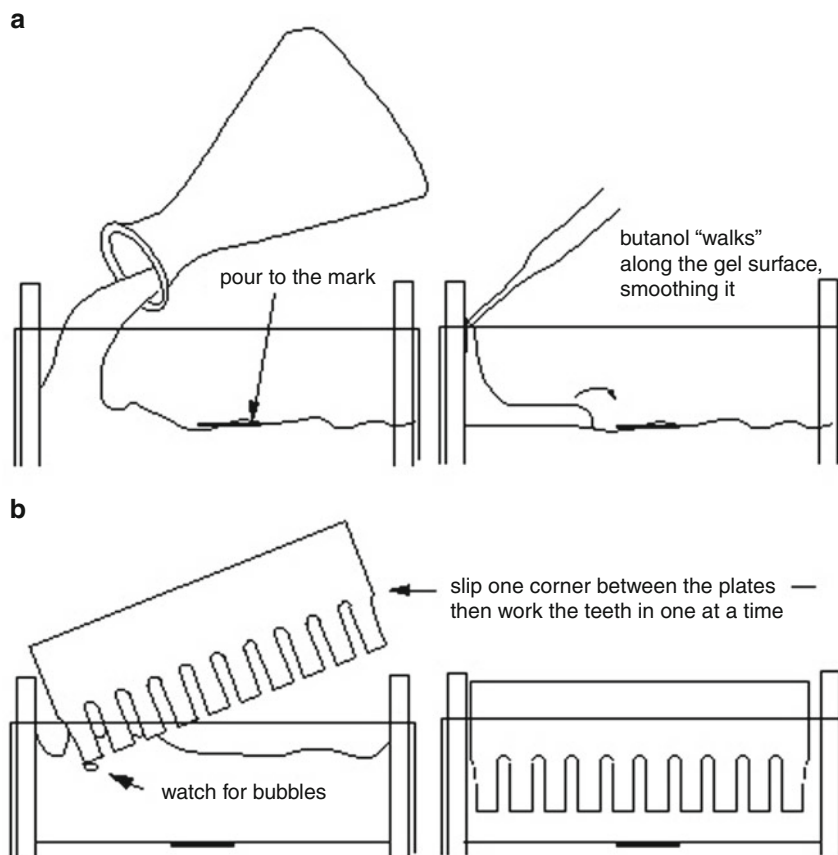


Fig. 5.4 (a, b) Preparation of casting gel

into the pipet, allowing it to stand, and checking it after 10 min or so. It should not take more than 15 minutes for any of the gel mixes to polymerize (Fig. 5.4).

5.3.2.4. Stacking Gel Preparation

Ten ml of stacking gel mix is sufficient for three of our cassettes; however, for the sake of accuracy it may be preferable to make 20 or 30 ml. Stacking gel buffer stock consists of 0.5 M Tris-Cl, pH 6.8, with 0.4 % SDS. Typical stackers are 3–4.5 % acrylamide. Before adding the final two components, which will start polymerization, the butanol should be poured off the separating gels into a sink with tap water running and excess butanol/acrylamide removed from the surfaces with a pipet. After adding AP and TEMED immediately swirl the mix and pour it into the cassettes to the tops of the plates. Insert combs one at a time, taking care not to catch bubbles under the teeth, and adjust to make them even if necessary, scraping excess stacking mix off later.

5.3.2.5. Preparing Protein Samples for Electrophoresis

A polypeptide is a macromolecule consisting of a nonbranching sequence of amino acids, each connected to the next by a single peptide bond. A protein consists of one or more polypeptides and/or additional types of molecules, held together by any of a number of molecular interactions often including covalent bonds. Such interactions result in several levels of organization, which we call primary, secondary, tertiary, and quaternary structures. Patterns of bands vary depending on temperature, buffer, variations in pH, quality of a preparation, etc. To characterize a type of preparation and obtain predictable results, we try to take proteins apart so that what we have left is primary structure only. The amino acid sequence of a polypeptide is called its *primary structure*. Interaction of soluble proteins with water leads to hydrogen bonding, which is partially responsible for the *secondary structure* of proteins. Secondary structure refers to the local structure of a polypeptide chain, including helices, pleated sheets, and turns. A functional protein has a three-dimensional structure resulting from hydrogen bonding, hydrophobic amino acids, Van der Waal's forces, and disulfide bonding. Three-dimensional structure of a protein is called its *tertiary structure*. Quaternary structure refers to the interaction of individual polypeptide chains with other molecules to form functional proteins. Although some proteins do consist of single polypeptides, many consist of two or more polypeptides linked by covalent bonds and/or noncovalent forces. In fact, many native (functional) proteins include non-protein components such as the carbohydrate groups on many membrane-associated proteins (Fig. 5.5a).

5.3.2.6. Sample Denaturation

Various sample buffers have been used for SDS-PAGE but all use the same principles to denature samples. Good denaturation by preparing a sample to a final concentration of 2 mg/ml protein with 1 % SDS, 10 % glycerol, 10 mM Tris-Cl, pH 6.8, 1 mM EDTA, a reducing agent such as dithiothreitol (DTT) or 2-mercaptoethanol, and a pinch of bromophenol blue to serve as a tracking dye (~0.05 mg/ml) (Fig. 5.5b, c).

2× concentrate of sample buffer consisting of 2 % SDS, 20 % glycerol, 20 mM Tris-Cl, pH 6.8, 2 mM EDTA, 160 mM dithiothreitol (DTT), and 0.1 mg/ml bromophenol blue dye.

What do the various components do?

1. EDTA is a preservative that chelates divalent cations, which reduces the activity of proteolytic enzymes that require calcium and magnesium ions as cofactors.
2. The tris acts as a buffer, which is very important since the stacking process in discontinuous electrophoresis requires a specific pH.

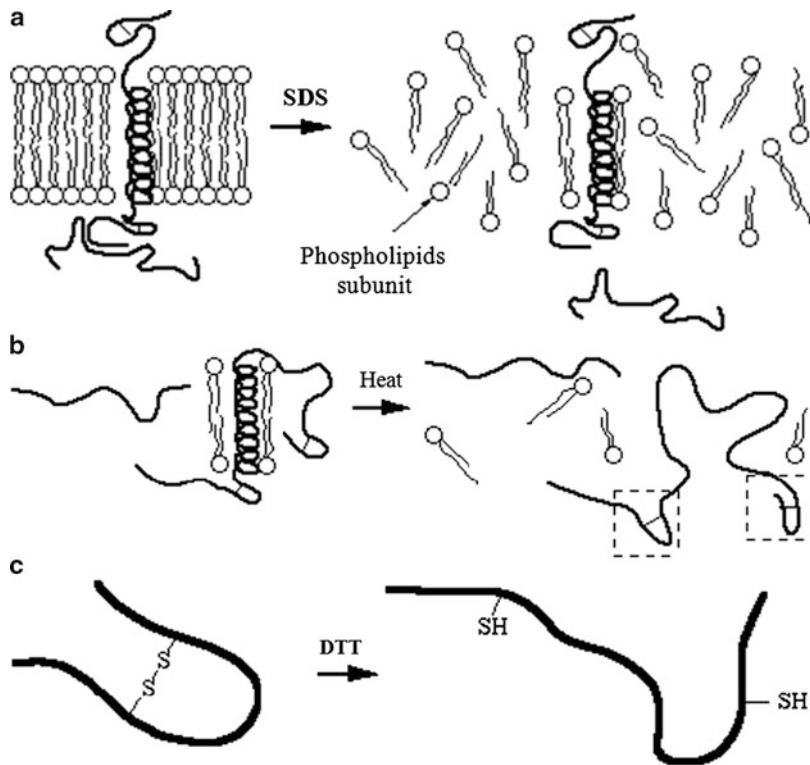


Fig. 5.5 (a) A native (functional) integral membrane protein is embedded in the phospholipid bilayer at *left*; at *right*, the anionic detergent has partially disrupted the interaction of protein and phospholipids. (b) Heating a sample in the presence of SDS speeds up the disruption of secondary, tertiary, and quaternary structure. *Dashed squares* indicate folds caused by disulfide bonds. Since they are covalent, disulfide bonds are not affected by SDS. (c) Dithiothreitol (DTT) reduces disulfide bonds, removing the last traces of tertiary or quaternary structure

3. Glycerol makes the sample more dense than the sample buffer, so the sample will remain in the bottom of a well rather than float out.
4. The dye allows the investigator to track the progress of the electrophoresis.
5. SDS breaks up the two- and three-dimensional structure of the proteins by adding negative charge to the amino acids, immediately rendering them functionless (Fig. 5a).
6. Heating the samples to at least 60 ° C shakes up the molecules, allowing SDS to bind in the hydrophobic regions and complete the denaturation (Fig. 5b).
7. DTT is a strong reducing agent. Its specific role in sample denaturation is to remove the last bit of tertiary and quaternary structure by reducing disulfide bonds (Fig. 5c).

5.3.2.7. Amounts to load

Polyacrylamide has a limited capacity for protein. Overloading results in precipitation and aggregation of proteins, producing

streaks and smears. Underloading results in complete disappointment. The objectives of sample preparation are to put the proteins into a denaturing buffer, rendering them suitable for electrophoresis, and to adjust the concentrations of sample so that an appropriate amount of protein can be loaded onto a gel.

The best results if we load 10 μ l of a 2 mg/ml final concentration of denatured protein per sample well. Dilute all samples to a predetermined concentration and volume before mixing with the denaturing buffer.

5.3.2.8. Assembling, Loading, and Running Gels

The assembly of a gel running stand varies with the type of apparatus. The top of the cassette must be continuous with an upper buffer chamber and the bottom must be continuous with a lower chamber so that current will run through the gel itself. The cassette must be sealed in place using gaskets or a sealant such as agarose. Fill both the upper and lower buffer compartments with an electrode buffer (running buffer) consisting of 25 mM Tris, 192 mM glycine, 0.1 % sodium dodecyl sulfate.

5.3.2.9. Loading Gels

Hamilton syringes work well for loading samples into the wells. Ideally, the glycerol in a sample causes it to sink neatly to the bottom of the well, allowing as much as 20 μ l or even more to be loaded.

5.3.2.10. Running gels

The anode (+ electrode) must be connected to the bottom chamber and the cathode to the top chamber. The negatively charged proteins will move toward the anode, of course. Gels are usually run at a voltage that will run the tracking dye to the bottom as quickly as possible without overheating the gels. Overheating can distort the acrylamide or even crack the plates. The voltage to be used is determined empirically. Run gels at 150 V.

5.3.2.11. Disassembly and Staining

When the dye front is nearly at the bottom of the gel, it is time to stop the run. For low percent gels with a tight dye front, the dye should be on the verge of running off the gel. When the percent of acrylamide is high the dye front may be diffuse, since the dye is not homogeneous. The plates are separated and the gel is dropped into a staining dish containing deionized water. After a quick rinse, the water is poured off and stain added. Staining usually requires incubation overnight, with agitation.

5.3.2.12. Staining Protein Gels

A commonly used stain for detecting proteins in polyacrylamide gels is 0.1 % Coomassie Blue dye in 50 % methanol and 10 % glacial acetic acid. Acidified methanol precipitates the proteins. Staining is usually done overnight with agitation. The agitation circulates the dye, facilitating penetration, and helps ensure uniformity of staining.

The dye actually penetrates the entire gel; however, it only sticks permanently to the proteins. Excess dye is washed out by 'destaining' with acetic acid/methanol, also with agitation. It is most efficient to destain in two steps, starting with 50 % methanol and 10 % acetic acid for 1–2 h, then using 7 % methanol and 10 % acetic methanol to finish. The first solution shrinks the gel, squeezing out much of the liquid component, and the gel swells and clears in the second solution. Properly stained/destained gels should display a pattern of blue protein bands against a clear background. The gels can be dried down or photographed for later analysis and documentation.

The original dye front, consisting of bromophenol blue dye, disappears during the process. In fact, bromophenol blue is a pH indicator which turns light yellow under acid conditions, prior to being washed out. In low percentage gels, sufficient protein may run with the dye front so that the position of the bromophenol blue front is permanently marked with unresolved proteins, often forming a continuous "front" across the bottom of the gel. In higher percent gels, a distinct dye front is usually not obtained.

Coomassie blue may not stain some proteins, especially those with high carbohydrate content. Stains such as periodic acid-Schiff (PAS), fast green, or Kodak 'Stain's all' may detect different patterns. Silver staining is generally used when detection of very faint proteins is necessary.

5.3.3. Silver Staining SDS PAGE Gels

1. Make 7 % acetic acid: 186 ml of water + 14 ml of acetic acid.
2. Make 50 % methanol: 200 ml of water + 200 ml of methanol.
Optional—for extra fixation/cross-linking add 240 μ l of 50 % glutaraldehyde to the 50 % methanol (makes solution 0.03 % glutaraldehyde).
3. Soak gel in 7 % acetic acid for 7 min.
4. Soak gel in 200 ml of 50 % methanol for 20 min for two times.
5. Prepare Solution A: 0.8 g of silver nitrate + 4 ml of water.
6. Rinse gel in ~200 ml water for 10 min

Note: Steps 7 and 8 are very important for the NuPAGE gels if you skip these steps or do not rinse the gel for long enough the gel will develop too quickly and have significantly more background.

7. 5 min before end of final water rinse prepare solution B: 21 ml of water + 250 μ l of 30 % NaOH (to make 0.36 %) + 1.4 ml of 14.8 M ammonium hydroxide.
8. To make staining solution: add solution A to solution B dropwise while stirring then add 76 ml of water.
9. Soak gel in the staining solution for 15 min.
10. Rinse gel in ~200 ml water for 5 min.

11. Rinse gel in ~200 ml water for 5 min.
12. Make developing solution: 200 ml of water + 1 ml of 1 % citric acid + 100 μ l of 37 % formaldehyde
13. Soak gel in developing solution until bands are visible usually 2–15 min
14. Stop development by rinsing gel with 3 changes of ~200 ml water
15. The sensitivity of this method should be in the 10 ng/band range.

References

1. Sambrook J, Russel DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, New York, NY
2. Berg JM, Tymoczko JL, Stryer L (2002) Biochemistry, 5th edn. WH Freeman, New York, NY, <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?&rid=stryer.section.438#455>
3. Robyt JF, White BJ (1990) Biochemical techniques theory and practice. Waveland Press, Illinois, IL
4. Agarose gel electrophoresis (basic method). Biological Protocols. <http://www.method-book.net/dna/agarogel.html>
5. Smisek DL, Hoagland DA (1989) Agarose gel electrophoresis of high molecular weight, synthetic polyelectrolytes. *Macromolecules* 22:2270
6. Schägger H (2006) Tricine-SDS-PAGE. *Nat Protocols* 1:16–22
7. Brody JR, Kern SE (2004) History and principles of conductive media for standard DNA electrophoresis. *Anal Biochem* 333:1–13
8. Bassam BJ, Caetano-Anolles G, Gresshoff PM (1991) Fast taxand sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem* 196:80–83

Molecular Identification of Microbes:

I. *Macrophomina phaseolina*

Bandamaravuri Kishore Babu, T. Kiran Babu, and Rajan Sharma

Abstract

This chapter will help us in the isolation of *Macrophomina phaseolina* from soil and infected plants and examination of morphological and physiological features for identification by using microscopic and cultural characters. In the later part, we will learn recent research findings to identify this fungus using PCR-based molecular techniques.

6.1 Introduction

Macrophomina phaseolina (Tassi) Goid., a soil-borne fungus, causes charcoal rot [1]. The fungus can infect the root and lower stem of over 500 plant species and is widely distributed all over the world. The pathogen causes wide range of diseases in the arid and semi-arid regions of the world. *M. phaseolina* persists in soil as sclerotia formed in infected host tissue and later released in the soil during decaying process. As a root inhabitant, the fungus is widespread in warmer area, invades immature, damaged, or senescent tissues; plants are generally attacked at seedling and flowering, when conditions are hot and dry. Infection develops from sclerotia, which can survive for a few years in roots. *M. phaseolina* is widely distributed among areas with variable soil types and annual rainfall, indicating that this fungus can persist under highly diverse environmental conditions. In case of soil-borne phase, the pathogen remains either on the dead organic debris or on the root stubbles, which are left over after the crop harvest. High soil temperature (40 °C), low soil pH (5.4–6), low soil moisture (8–16 %), and high humidity (90 %) favor infection and disease development. Long periods of drought and hot temperatures interspersed with rain showers create ideal conditions for the fungal pathogenesis.

Table 6.1
List of PCR primers

Primer name	Sequence	PCR product size
Universal primers		
ITS-1	5' TCCGTAGGTGAACCTGCGG 3'	~650
ITS-4	5' TCCTCCGCTTATTGA TATGC 3'	
Specific primers		
MPKF1	5' CCGCCAGAGGACTATCAAAC 3'	~350
MPKR1	5' CGTCCGAAGCGAGGTGTATT 3'	

6.2 Materials

- Diseased field soil
- Mesh-2 mm, 45 µm
- Sodium hypochlorite
- Acidified Potato Dextrose agar plates (pH 5.6)
- *M. phaseolina* culture
- Potato Dextrose broth
- Lysis buffer—(50 mM Tris-HCl, pH 7.8, 50 mM Na₂-EDTA, 3 % SDS), 1 % 2-mercaptoethanol should be added freshly)
- Primers (Table 6.1)
- Thermal cycler
- *Taq* DNA polymerase
- 10× PCR buffer
- Milli Q Water
- 50× TAE buffer

6.3 Methods

6.3.1. Isolation

1. Sieve the air-dried soil through mesh.
2. Dissolve 5 g of soil in 0.525 % sodium hypochlorite and allow standing for 10–20 min.
3. Wash the deposit in sterile distilled water over a sieve with a 45 µm mesh.

4. Introduce the deposit into a 250-ml flask and incorporate in to 100 ml of PDA.
5. Pour into the petriplates and incubate at 32–34 °C for 3–4 days.
6. Colony morphology: On PDA colonies range in color from white to brown or gray and darken with age.

6.3.2. Microscopy

1. Hyphal branches generally form at right angles to parent hyphae, but branching is also common at acute angles.
2. Pycnida: 100–200 µm in diameter; dark to grayish, becoming black with age; globose or flattened globose; membranous to subcarbonaceous with an inconspicuous or definite truncate ostiole.
3. The pycnida bear simple, rod-shaped conidiophores, 10–15 µm long.
4. Conidia: 14–33 × 6–12 µm, single celled, hyaline, and elliptic or oval.
5. Microsclerotia: jet black in color and appear smooth and round to oblong or irregular.

6.3.3. DNA Extraction

See Chap. 1 (Fungal DNA isolation).

6.3.4. Polymerase Chain Reaction

Polymerase chain reaction (PCR) is a technique used to amplify a part of DNA that lies between two regions of known sequences. Two oligonucleotides are used as primers for a series of synthetic reactions that are catalyzed by DNA polymerase. These oligonucleotides specifically anneal to the target sequences on opposite strands and flank the region that is to be amplified.

6.3.4.1. Amplification of 5.8S rDNA Gene and ITS Regions

Prepare the PCR master mix following Table 6.2. Redistribute 44 µl of master mix into PCR tube and add 4–6 µl genomic DNA.

6.3.4.2. PCR Program for ITS Primers

Lid heating—Enabled	
Step 1 = 95 °C for 5 min	(Initial denaturation)
Step 2 = 95 °C for 1 min	(Denaturation)
Step 3 = 50 °C for 30 s	(Primer annealing)
Step 4 = 72 °C for 1 min 20 s	(Elongation)
Step 5 = 34 Cycle	(Repeat steps 2–3)
Step 6 = 72 °C for 10 min	(Final elongation)
Hold at 4 °C for 10 min	

Table 6.2
Preparation of PCR master mixture for single reaction

Reaction mixture	Primer sets and reagents concentration	
	ITS1 and ITS4	MPKF1 and MPKR1
Genomic DNA	20–40 ng	10–25 ng
Forward primer	50 pmol	5 pmol
Reverse primer	50 pmol	5 pmol
dNTPs mix	0.2 mM	0.2 mM
10× PCR buffer	5 µl	2 µl
<i>Taq</i> DNA polymerase	1 U	0.4 U
Milli Q Water	Make up to 50 µl	Make up to 20 µl

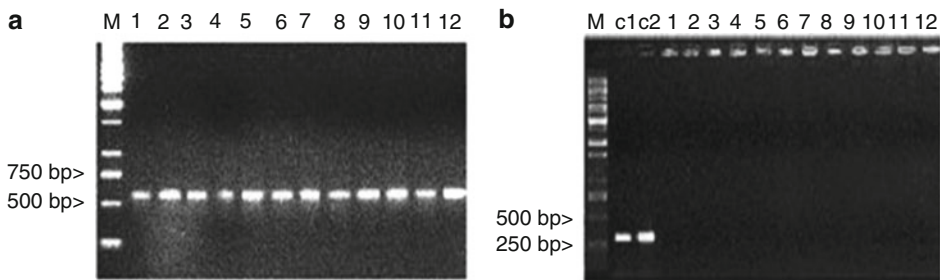


Fig. 6.1 PCR Amplification of rDNA gene cluster: (a) Primers ITS1 and ITS4 were used for amplification of nearly 650 bp fragment. Lanes 1–10 showing amplified products of *M. phaseolina* isolates. (b) Amplification of *M. phaseolina* with specific primers (*MpKF1* and *MpKR1*) produced 350 bp amplicon in lanes c1 and c2. Lanes 1–12 showing no amplified product with different test microbes. M- 1 kb molecular ladder.

6.3.4.3. PCR Program for
***M. phaseolina* Specific**
Primers [1]

Lid heating—Enabled	
Step 1 = 95 °C for 5 min	(Initial denaturation)
Step 2 = 95 °C for 30 s	(Denaturation)
Step 3 = 56 °C for 1 min	(Primer annealing)
Step 4 = 72 °C for 2 min	(Elongation)
Step 5 = 25 Cycle	(Repeat steps 1–3)
Step 6 = 72 °C for 10 min	(Final elongation)
Hold at 4 °C for 10 min	

6.3.5. Gel Electrophoresis

PCR amplified products together with marker (1 kb Fermentas, USA) were resolved by gel electrophoresis (4 V cm^{-1}) on 1.4 % agarose gels in $1 \times$ TAE buffer containing 0.5 mg ml^{-1} Et-Br and visualized under UV transilluminator (Fig. 6.1).

Reference

1. Babu KB, Srivastava AK, Saxena AK, Arora DK (2007) Identification and detection of *Macrophomina phaseolina* by using species-specific oligonucleotide primers and probe. Mycologia 99:733–739

Chapter 7

Molecular Identification of Microbes: II. *Bacillus*

Anil Kumar Saxena

Abstract

Comparison of 16S rDNA sequences from the type strains of 69 species revealed that the 5' end region was the hypervariant region (HV region) and highly specific for each type strain. Further the HV region is highly conserved among the species. Thus the sequencing of HV region is a very efficient index for the rapid identification or grouping of *Bacillus* species. A simple procedure for identification of genus *Bacillus* per se and to identify species based on sequencing of only a small fragment of 16S rRNA was performed. The comparative analysis of restriction site for *AluI* in 16S rRNA gene sequence could identify a 265 bp band common to at least 25 species but absent in newly created *Bacillus*-related genera and phenotypically related genera. Primers were designed to amplify the hypervariant region in the 265 bp fragment and the sequencing of this small region was found as a useful criterion to classify species of *Bacillus*.

7.1 Introduction

The genus *Bacillus* is a large, heterogeneous group of Gram-positive, aerobic, endospore-forming, rod-shaped bacteria. Since endospore formation is a universal feature of these bacteria, spore morphology has traditionally been given considerable weightage in their classification and identification. The genus *Bacillus* was established by Cohn [1]. Since then it has undergone considerable taxonomic changes. It started with two prominent and truly endospore-forming species, *Bacillus anthracis* and *Bacillus subtilis* and their number increased to an incredible 146 in the fifth edition of Bergey's Manual of Determinative Bacteriology. In the sixth edition it reduced to 33, in the seventh to 25, and in the eighth edition to 22 well-defined species and 26 species that received less recognition.

With the introduction of modern taxonomic techniques such as numerical phenetics, DNA base composition determinations, and DNA reassociation experiments, DNA sequence homology between strains can be estimated; it became apparent that the

bacilli were more heterogeneous than hitherto suspected. The range of DNA base composition among strains is a good indicator of genetic diversity; it is generally agreed that species in a genus should not vary by more than 10–12 mol% G+C. In the case of *Bacillus*, the range is about 33–65 % although strains of most species cluster between 40 % and 50 % [2]. Different studies have grouped the species into various clusters depending upon the characters used [3–5]. Studies based on comparative analysis of 16S rDNA gene sequences of different *Bacillus* species revealed five phylogenetically distinct clusters [3]. Further characterization at the genotypic and phenotypic levels of selected *Bacillus* species have led to the creation of several new genera like *Alicyclobacillus*, *Paenibacillus*, *Brevibacillus*, *Virgibacillus*, *Geobacillus*, *Filobacillus*, *Jeotgalibacillus*, *Aneurinibacillus*, *Gracibacillus*, and *Marinibacillus*. The *Bacillus* species along with the species from related genera were identified successfully and differentiated by rRNA gene restriction patterns, and three distinct main genetic clusters at the 75 % banding pattern similarity were obtained [6]. In general 16S rDNA sequences are used in *Bacillus* classification as a framework of species delineation [7]. Today over 200 species of aerobic, endospore-forming bacteria (AEFB) allocated to about 25 genera have been validly published.

7.2 Methods for the Identification of *Bacillus* and *Bacillus*-Derived Genera

7.2.1. Phylogenetic Relationship Between *Bacillus* Species and Related Genera Through Comparison of 3' End of the 16S rDNA and 5' End of the 16S–23S ITS Nucleotide Sequences

Ribosomal RNA sequences were being established as the most useful molecular chronometer to infer phylogenetic relationships because they are present in all organisms and changes in the nucleotide sequences were deemed to occur in a clocklike manner [8]. A comparison of 16S and 16–23S ITS nucleotide sequences led to the identification of conserved regions in almost 46 different *Bacillus* species. A primer pair was developed from the conserved region; one located about 200 nt upstream from the 3' end of the 16S rRNA gene, the other about 80 nt downstream from the 5' end of the 23S rRNA gene. It could amplify the last 200 bp of the 16S rRNA gene and the entire 16S–23S ITS region. The amplified fragments vary in length from 450 to 850 bp [5]. Sequencing of the amplified product could be used to establish the phylogenetic relationship among the *Bacillus* species.

The protocol is as follows:

1. Grow the *Bacillus* strains in Nutrient broth at 30 °C for 24–48 h to O.D of about 0.6.
2. Isolate the genomic DNA from the culture following the protocol described in this book for Gram positive bacteria.
3. Amplify the 3' end of 16S rDNA, the 16S–23S ITS region, and the 5' end of 23S rDNA with a pair of primers: L516SF (5'-TCGCTAGTAATCGCGGATCAGC-3') and L523SR (5'-GCATATCGGTGTTAGTCCCCGTCC-3').
4. Perform amplification in a total volume of 50 µl containing about 50 ng DNA, 0.25 µM each primer, 200 µM dNTP, 1.5 mM MgCl₂, and 1.25 U of *Taq* DNA polymerase.
5. Perform PCR under the following conditions: 45 s at 95 °C and then 30 cycles of 15 s at 94 °C, 30 s at 53 °C, and 90 s at 72 °C.
6. View the amplification products on 0.8 % agarose gels.
7. Clone the amplified DNAs into a pCRII-TOPO cloning vector using the TOPO TA cloning kit, following the manufacturer's instructions.
8. Sequence the cloned fragments using an automated DNA sequencer.
9. Align the sequences using the CLUSTAL W program [9] and construct the most parsimonious phylogenetic trees using the DNAPARS program of the PHYLIP package, version 3.6a2 [10, 11].

7.2.2. Partial 16S rDNA Sequence for Rapid Identification of Species in the Genus *Bacillus*

Identification of *Bacillus* species based on the partial sequencing of 16S rDNA was described by Goto et al. [12]. Comparison of 16S rDNA sequences from the type strains of 69 species revealed that the 5' end region was the hypervariable region (HV region) and highly specific for each type strain. Further the HV region is highly conserved among the species. Thus the sequencing of HV region is a very efficient index for the rapid identification or grouping of *Bacillus* species.

The protocol is as follows:

1. Grow the *Bacillus* strains in Nutrient broth at 30 °C for 24–48 h to a O.D of about 0.6.
2. Isolate the genomic DNA from the culture following the protocol described in this manual for Gram positive bacteria.
3. Amplify the HV of 16S rDNA with a pair of primers: a forward primer: 5'-TGT AAA ACG ACG GCC AGT GCC TAA TAC ATG CAA GTC GAG CG-3' and a reverse primer: 5'-CAG GAA ACA GCT ATG ACC ACT GCT GCC TCCCCGT AGG AGT-3'.

4. Perform amplification in a total volume of 50 µl containing about 50 ng DNA, 100 ng of each primer, 200 µM dNTP, 1.5 mM MgCl₂, and 1.25 U *Taq* DNA polymerase.
5. Perform PCR under the following conditions: 5 min at 95 °C and then 30 cycles of 45 s at 94 °C, 30 s at 53 °C, and 90 s at 72 °C.
6. View the amplification products on 0.8 % agarose gels.
7. Perform the sequence analysis using the EMBL, GenBank, and DDBJ database and carry on BLAST search to look for nearest neighbor.

7.2.3. ARDRA and Partial Sequencing of rRNA as an Index to Identify *Bacillus* species

Several new genera like *Alicyclobacillus*, *Paenibacillus*, *Brevibacillus*, *Virgibacillus*, *Geobacillus*, *Filobacillus*, *Jeotgalibacillus*, *Aneurinibacillus*, *Gracibacillus*, and *Marinibacillus* have been derived from the genus *Bacillus*. A simple procedure for identification of genus *Bacillus* per se and to identify species based on sequencing of only a small fragment of 16S rRNA was performed. The comparative analysis of restriction site for *AluI* in 16S rRNA gene sequence could identify a 265 bp band common to at least 25 species but absent in newly created *Bacillus*-related genera and phenotypically related genera. Primers were designed to amplify the hypervariable region in the 265 bp fragment and the sequencing of this small region was found as a useful criterion to classify species of *Bacillus*.

The protocol is as follows:

1. Grow the *Bacillus* strains in Nutrient broth at 30 °C for 24–48 h to a O.D of about 0.6.
2. Isolate the genomic DNA from the culture following the protocol described in this book for Gram positive bacteria.
3. Amplify 16S rDNA with 60–100 ng of pure genomic DNA using the forward (PA) 5'-(AGAGTTTGATCCTGGCTCAG)-3' and reverse (PH) 5'-(AAGGAGGTGATCCAGCCGCA)-3' primers.
4. Perform amplification reaction in a 100 µl volume by mixing template DNA with the polymerase reaction buffer (1×); 80 µM (each) dATP, dCTP, dTTP, and dGTP; primers PA and PH (100 ng each); and 1.5 U *Taq* polymerase.
5. Perform PCR amplification with the following temperature profile: an initial denaturation for 30 s at 94 °C for 5 min, 30 cycles of denaturation for 30 s at 94 °C, annealing at 52 °C for 40 s, and extension at 72 °C for 1 min 30 s, and a final extension at 72 °C for 7 min.



Fig. 7.1 Amplification of 220 bp fragment of 16S rDNA using the primer pair 265 F1 and 265R1.

6. Run the amplified product on a 0.8 % agarose gel along with λ *Hind*III molecular weight marker at a constant voltage and visualize the gel following staining with ethidium bromide in gel documentation system.
7. Digest an aliquot of purified 16S rDNA with restriction endonuclease *Alu*I in 25 μ l reaction volume by using the manufacturer's recommended buffer and temperature.
8. Run the restriction products in 2.5 % (w/v) agarose gel in TE buffer for 3 h. Stain the gel and view as described above.
9. Look for the presence of 265 bp fragment. If present, the species belongs to *Bacillus* with few exceptions like *B. cereus*, *B. thuringiensis*, *B. anthracis*, and *B. mycoides*. If the fragment is absent, it may be a *Bacillus*-derived genera.
10. Perform another PCR amplification reaction to amplify only the 265 bp fragment using a primer pair: 265F1: 5'-GTGCTACAATGGACAGAACAA-3' and 256R1: 5'-GTGAGATGTTGGGTTAAGTC-3' using the reaction and amplification conditions as described above for amplification of 16 S rDNA. The primers will amplify a product of 220 bp (Fig. 7.1).
11. Gel extract the fragment using commercially available gel extraction kit or purify the fragment using PCR purification kit.
12. Sequence the product and BLAST search for nearest neighbor using EMBL, GenBank, and DDBJ database.

References

1. Cohn F (1872) Untersuchungen u"ber Bakterien. Beitr Biol Pflanz 1:127–224 (in German)
2. Priest FG (1981) DNA homology in the genus *Bacillus*. In: Berkeley RCW, Goodfellow M (eds) The aerobic endospore-forming bacteria: classification and identification. Academic Press, Inc., London, pp 33–57
3. Ash C, Farrow AE, Wallbanks S, Collins MD (1991) Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. Lett Appl Microbiol 13:202–206
4. Priest FG, Goodfellow M, Todd C (1988) A numerical classification of the genus *Bacillus*. J Gen Microbiol 134:1847–1882
5. Xu D, Cote JC (2003) Phylogenetic relationships between *Bacillus* species and related genera inferred from comparison of 39 end 16S rDNA and 59 end 16S–23S ITS nucleotide sequences. Int J Syst Evol Microbiol 53: 695–704
6. Joung KB, Cote JC (2002) Evaluation of ribosomal RNA gene restriction patterns for the classification of *Bacillus* species and related genera. J Appl Microbiol 92:97–108
7. Stackebrandt E, Swiderski J (2002) From phylogeny to systematics: the dissection of the genus *Bacillus*. In: Berkeley R, Heyndrickx M, Logan NA, De Vos P (eds) Applications and systematics of *Bacillus* and relatives. Blackwell Publishing, Oxford, United Kingdom

8. Woese C (1987) Bacterial evolution. *Microbiol Rev* 51:221–271
9. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
10. Felsenstein J (1989) PHYLIP-phylogeny inference package (version 2). *Cladistics* 5:164–166
11. Felsenstein J (2001) PHYLIP version 3.6a2. Distributed by the author. Department of Genetics, University of Washington, Seattle, WA, USA.
12. Goto K, Omura T, Hara Y, Sadaie Y (2000) Application of the partial 16S rDNA sequence as an index for rapid identification of species in the genus *Bacillus*. *J Gen Appl Microbiol* 46: 1–8

Chapter 8

Molecular Identification of Microbes: III. *Pseudomonas*

Bhim Pratap Singh and Ratul Saikia

Abstract

The molecular identification protocol for *Pseudomonas* introduces PCR cycle sequencing and some basic bioinformatics tools. First follow a simple protocol to isolate genomic DNA from the bacteria. This protocol involves breaking the cells open with a series of freeze/thaw cycles and then centrifuging to remove cellular debris. Many techniques exist for the identification of *Pseudomonas*. We present a detailed explanation for setup of 16S rRNA amplification and 16S rRNA RFLP. First after isolation of DNA a PCR reaction has to set up to amplify a region of the 16S rRNA gene. The PCR product should be cleaned up by using PCR purification kit, which cleaves excess primers and inactivates free nucleotides. The cleaned PCR product is then used as the template for a sequencing reaction. Sequencing should be done by using BigDye reagents and the reactions are run in a thermocycler (PCR machine). The completed samples are then sent to a core facility to obtain the sequence. Finally, view the electropherograms from the sequencing reaction, then use the sequence in a BLAST search limited to a bacterial data base. Students can identify their unknown bacteria by examining the top-scoring sequences from the BLAST search results. 16S rRNA-RFLP is an important tool to understand the phylogenetic similarity between isolates so that one can avoid the sequencing of the isolates which are similar and it leads to the reduction of cost.

8.1 Introduction

Pseudomonas is a member of the Gammaproteobacteria class of Eubacteria. It is a Gram-negative, free-living aerobic rod belonging to the family Pseudomonadaceae commonly found in soil and water. Since the revisionist taxonomy based on conserved macromolecules (e.g., 16S ribosomal RNA) the family includes only members of the genus *Pseudomonas* which are cleaved into eight groups; the best studied species include *P. aeruginosa* in its role as plant growth promoting *P. fluorescens* [1–3], plant pathogen *P. syringae* [4], an opportunistic human pathogen [5]. Their ease of culture in vitro and availability of an increasing number

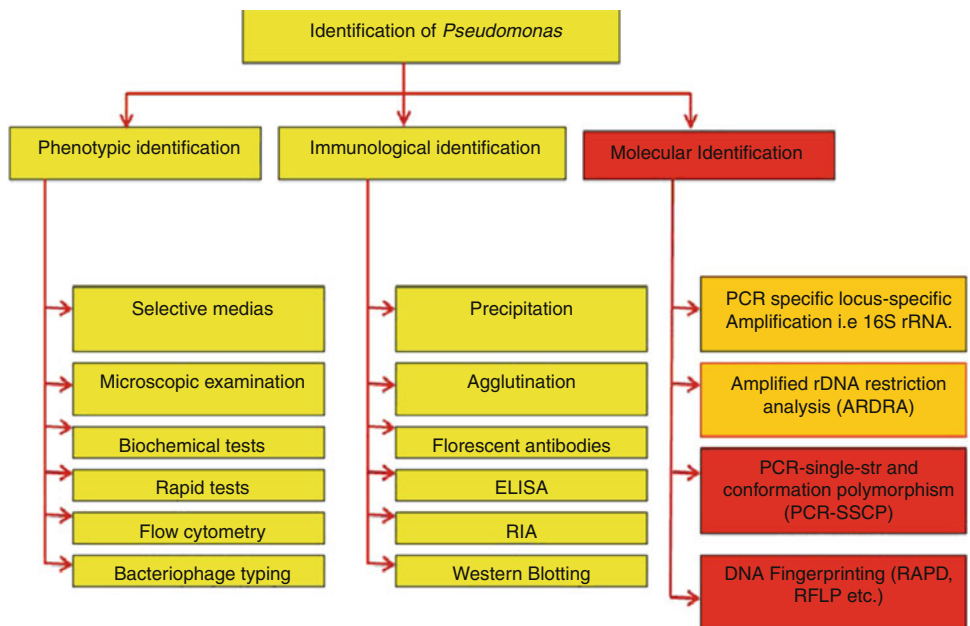


Fig. 8.1 Various methods used for the identification of *Pseudomonas* spp.

of *Pseudomonas* strain genome sequences has made the genus an excellent focus for scientific research.

There is increasing interest in the use of molecular methods by the application of polymerase chain reaction (PCR) technology for the identification of microbial community. Such methods offer the advantage of reducing or eliminating the need for lengthy culturing and difficult morphological identification procedures (Fig. 8.1).

PCR is a very simple reaction in both concept and practice. First, template DNA that contains some region of interest is isolated. The template does not need to be purified, and one can start with a very tiny amount of template. Here, our source of template DNA will be the genome of your unknown bacterial strain. The experimenter does not need to have some prior sequence information about the segment to be amplified, because primers (for DNA replication) need to be used that are complementary to sequence on both sides of the segment to be amplified. In this experiment, we will be using “universal” 16S rRNA gene primers. Because we are hoping to amplify genes that are similar but not identical from the different bacteria, primers need to be designed to anneal to the parts of the DNA sequence that are most similar in all bacteria. The actual production of DNA is carried out

by a DNA polymerase enzyme, which starts at the primers and uses the template to produce copies of the template sequence. Many copies of the template are created by repeating the reaction many times (i.e., many cycles).

After creating numerous copies of the 16S rRNA genes from unknown bacteria, we will determine the sequence of the DNA using a modified “dideoxy” sequencing reaction. Once the DNA sequence is determined, we will compare the sequence to known sequences in a computer database. The 16S rRNA genes of our bacterial species have already been sequenced, so that it is likely that the sequence of the 16S rRNA gene of your unknown is in the database. The potential benefits of this technology have been especially recognized in the regulatory field, where timeliness and accuracy of identifications are crucial. *Pseudomonas* is one of the important genus which has great potential exists in it, so we need to develop certain molecular protocols to identify this genus.

8.2 Materials

8.2.1. Amplification of 16S rRNA Gene

1. UV-VIS spectrophotometer: To measure concentration of DNA.
2. Universal primers: Universal primers for the amplification of 16S rRNA gene. The primers can also be synthesized commercially (see Note 1).
3. Polymerase chain reaction (PCR): A thermal cycler to amplify the specific region of DNA.
4. Gel electrophoresis Unit: Required to run the gel.
5. Gel documentation system: To view the gel photos (see Note 2).
6. Sequencing: It can be done commercially by different institutions.
7. Bioinformatics facility: Basic bioinformatics facility to analyze the sequenced data for the identification of bacteria.

8.2.2. 16S rRNA-RFLP

1. In addition to 2.1.
2. Restriction endonucleases, mainly tetra-cutters like *MspI*, *HaeIII*, *AluI*, etc. can be procured from any chemical supplying company.
3. Ntsys 2.02 software to analyze the phylogenetic relationship among the isolates.

8.3 Methods

8.3.1. Polymerase Chain Reaction Mediated Locus-Specific Amplification from the Genomic DNA of *Pseudomonas*

Microbial community analysis based on small-subunit rRNA genes as well as protein-coding gene clone libraries has become common practice in microbial ecology. The PCR has enabled specific genetic loci to be routinely amplified and examined for differences indicative of strain variation. The specific locus is examined and amplified with gene-specific primers and subjected to RFLP analysis. The DNA fragments are separated on an agarose gel, and the digested patterns are visualized following ethidium bromide staining.

Analysis of the 16S rRNA gene sequence is of fundamental importance to current prokaryote biodiversity studies and phylogenetic analyses. Most 16S rRNA gene sequences are generated through PCR amplification of mixed template samples and as a consequence an increasing number of chimeric 16S rRNA sequence records are being deposited into the public repositories [6]. Locus-specific RFLP has been applied in a number of situations. The 16S, 23S, and 16S–23S spacer regions have been used as targets for locus-specific RFLP. In this variation of ribotyping, the ribosomal DNA is amplified and subjected to digestion with restriction enzyme, and the DNA fragments are visualized following separation by gel electrophoresis. The experimental lay-out is described in Figs. 8.2 and 8.3.

8.3.2. Amplification of 16S rRNA

1. The genomic DNA isolated from unknown bacteria will be quantified spectrophotometrically by measuring the absorbance at 260 nm. The amount of DNA will be estimated using the relationship that OD of 1.0 corresponds to 50 $\mu\text{g ml}^{-1}$. Part of DNA samples should be diluted with appropriate amount of Milli-Q water to yield a working concentration of 50 ng/ μl and stored at 4 °C.
2. Primers: Universal primers can be used for 16S rRNA amplification. The stock solution (100 ng/ml) of the primers need to be prepared by reconstituting lyophilized primers in Milli-Q water and stored at –20 °C.
3. Amplification reactions can be performed in a 100 μl volume, containing: 1 \times PCR Buffer (4 μl), 2 mM dNTPs (4 μl), 100 ng of each primer (1.5 μl), 50 ng of template DNA (3 μl), 1 unit Taq DNA Polymerase (0.60 μl) and the volume was adjusted by Milli-Q water. The PCR protocol is as follows: initial denaturation at 94 °C for 5 min, 40 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 40 s, and extension at 72 °C for 1 min, and final extension at 72 °C for 10 min (see Note 3).

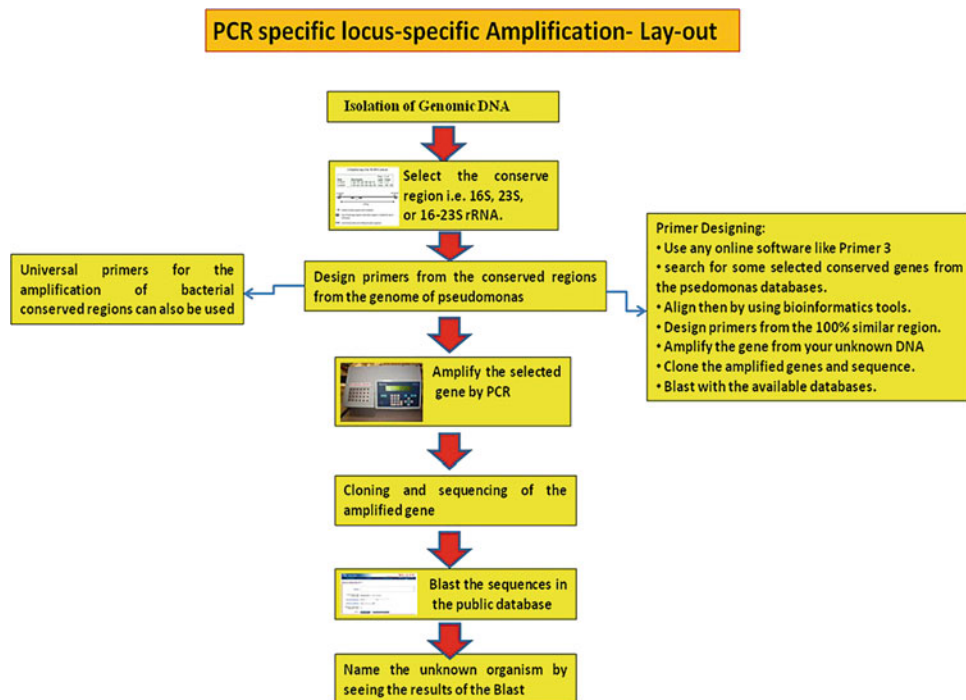


Fig. 8.2 Steps involved in the identification of pseudomonas by amplification of 16S rRNA gene.

4. Tris-Acetate-EDTA (TAE) buffer—50× stock solution. For running gels.
 - 242 g Tris base.
 - 57.1 ml glacial acetic acid.
 - 100 ml 0.5 M EDTA (pH 8.0).
 Make up the volume to 1 l by adding dd H₂O. Autoclave and use 1× in the running tray and gel preparation (see Note 4).
5. Ethidium bromide: Stock should be prepared by dissolving 10 mg in 1 ml of dd H₂O (see Note 5).
6. 6× gel loading dye: Add 0.25 % bromophenol blue; 0.25 % xylene cyanol; 30 % Glycerol; 60 mM EDTA. The volume was made up to 20 ml with millipore water.

8.3.3. 16S rRNA-RFLP

1. Amplify the 16S rRNA gene by using gene universal primers as mention above along with control (some known *Pseudomonas* spp.) following Fig. 8.4.
2. Select any one tetra-cutter restriction enzyme (*AluI*, *HaeIII*, or *MspI*).

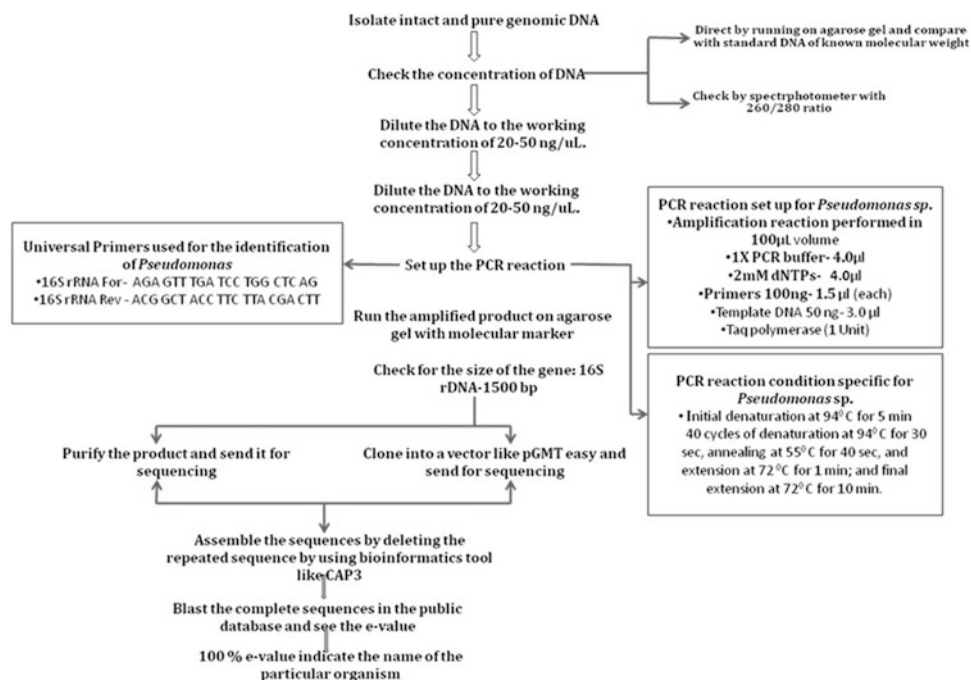


Fig. 8.3 General set up to perform the amplification of 16S rRNA gene in detail with experimental lay out for the identification of *Pseudomonas*.

3. Perform restriction digestion with any one enzyme-amplified DNA (amplified product of PCR)-100 ng, 10× enzyme specific buffer and 1 U of enzyme; mix in an eppendorf tube and incubate at 37 °C for 1 h.
4. Run the digested product for on 3 % agarose gel.
5. Record the banding pattern among all the isolates.
6. Construct the dendrogram by using NTsys 2.02 or any other available online or offline software (see Note 6).
7. Analyze the position of unknown isolates with respect to control (known *Pseudomonas*).
8. Send one or two representative isolates from each cluster (see Note 7).

8.4 Notes

1. The universal primers 16S rRNA. (F-AGA GTT TGA TCC TGG CTC AG and R-ACG GCT ACC TTC TTA CGA CTT) can be used for PCR amplification of 16S rDNA [5]. These

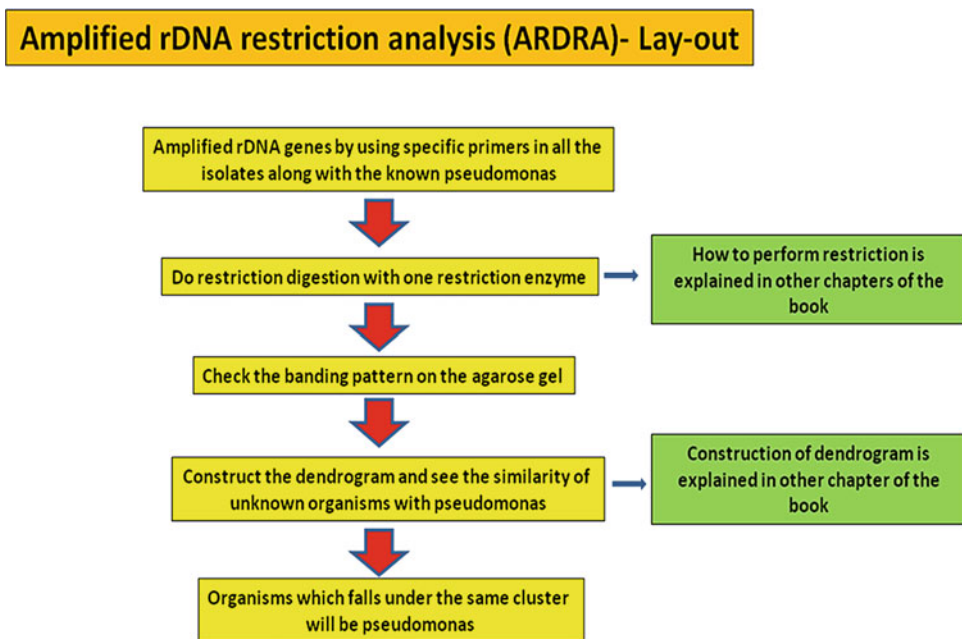


Fig. 8.4 Experimental lay-out to perform 16S rRNA RFLP.

primers can be synthesized or can be procured commercially. Primers should be diluted to the working concentration of 50–100 pmol/μL.

2. Visualize the gel carefully and do the scoring as “1” for presence of band and “0” for absence of band comparing to the known molecular weight marker.
3. If some problem persists with PCR conditions, it can be checked and rectified through <http://www.pcronline.com/>.
4. TAE buffer should be diluted to 1× of 50×. Autoclave the buffer before use. Always prepare the gel in 1× TAE buffer not dd H₂O. Prepare the EDTA separately with pH 8.0 and add into the buffer.
5. Ethidium Bromide is a carcinogenic chemical, always wear gloves before handling EtBr.
6. For dendrogram construction you can use many online (Dendrogram plot, hierarchical clustering, Phylip, etc.) as well as offline (BioEdit, NTSys, DNA star, etc.) softwares.
7. Sequencing can be done commercially from different institutions.

References

1. Kropp BR, Thomas E, Pounder JI, Anderson AJ (1996) Increased emergence of spring wheat after inoculation with *Pseudomonas chlororaphis* isolate 2E3 under field and laboratory conditions. *Biol Fert Soil* 23:200–206
2. Leisinger T, Margraff R (1979) Secondary metabolites of the fluorescent pseudomonads. *Microbiol Rev* 43:422–442
3. Saikia R, Varghese S, Singh BP, Arora DK (2009) Influence of mineral amendment on disease suppressive activity of *Pseudomonas fluorescens* to *Fusarium* wilt of chickpea. *Microbiol Res* 164:365–373
4. Jaunet T, Laguerre G, Lemanceau P, Frutos R, Nottoghem JL (1995) Diversity of *Pseudomonas fucovaginae* and other fluorescent pseudomonads isolated from diseased rice. *Phytopathology* 85:1534–1541
5. Palleroni NJ (1993) *Pseudomonas* classification: a new case history in the taxonomy of Gram-negative bacteria. *Antonie Van Leeuwenhoek* 64:231–251
6. Weisburg WG, Barns SM, Pelletier DA, Lane DJ (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 173:697–703

Chapter 9

Molecular Identification of Microbes: IV. *Vibrio*

Hirak Ranjan Dash, Neelam Mangwani, and Surajit Das

Abstract

Vibrios are ubiquitous and abundant in aquatic environment. A huge fraction of *vibrios* have also been detected in marine environments. Being of environmental origin, *Vibrio* species can cause the diseases like cholera and gastroenteritis. Thus detection of the species becomes utmost important when there is the need of disease diagnosis and the causative agent detection. Though there are many techniques available for the identification of *Vibrio* species manually, they are time-consuming procedures and not economical. Hence in this chapter, we have described the techniques of identification of *Vibrio* species at molecular level by house-keeping gene amplification and PCR fingerprinting techniques which will be economical as well as time saving for rapid identification to boost the diagnosis procedures.

9.1 Introduction

Vibrio spp. are curved, rod-shaped, gram negative bacteria found in brackish salt water, being the normal microflora of the marine environment. When ingested they cause gastrointestinal illness in human and are reported to be the most serious pathogens in fish and shell fish in marine aqua culture worldwide. They possess the unique characteristic feature of being oxidase positive, facultatively aerobic, and non-spore formers. They are motile, possessing single polar flagellum [1]. Most of the *Vibrio* species are pathogenic in nature, which are associated with gastroenteritis and septicemia include *V. cholera*, *V. parahaemolyticus*, *V. vulnificus*, *V. fischeri*, *V. harveyi* [2]. Almost all people are at great risk of developing disease to *Vibrio* species whether they are healthy people or immune-compromised people. Healthy hosts who ingest large quantities of *Vibrio* may experience only gastroenteritis where as immune-compromised patients having histories of liver disease are more vulnerable to septicemia.

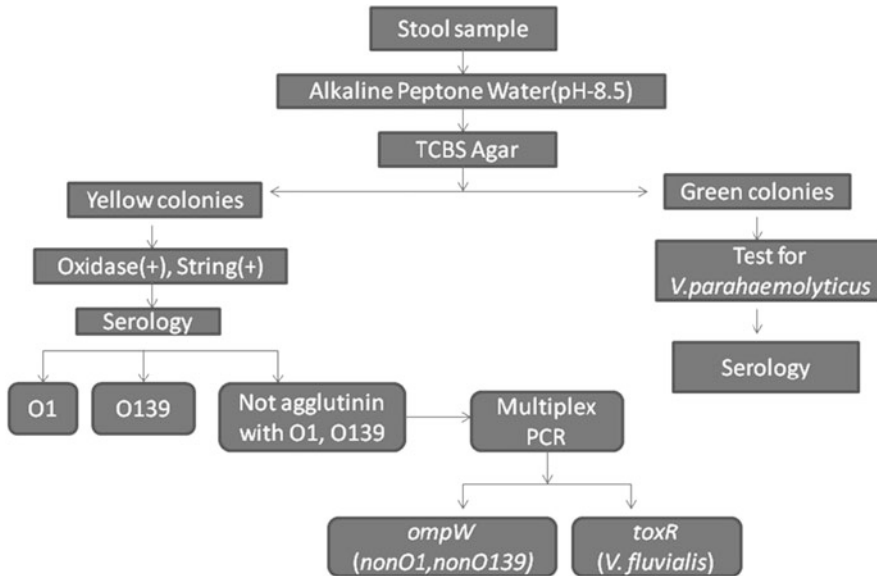


Fig. 9.1 Flow chart for identification of *Vibrio cholerae* and other species of *Vibrio*.

Vibrio infection accounts for 69 % of the total food-borne outbreaks in the globe which can become more severe if untreated, but when treated the rate of fatality due to *Vibrio* infection is reduced to less than 1 %. The treatment for *Vibrio* infections includes fluid and electrolyte replacements and antibiotic treatments. For targeted and quick treatment of bacterial infection, identification of the cause of infection is required at a rapid rate. Though various culture-based identification procedures are available presently for *Vibrio* species, they are time consuming and not so accurate. Molecular procedures involving identification of *Vibrio* species are quite rapid and accurate which are more efficient in terms of identification and feasibility. Currently molecular methods of identification are often used in addition to or instead of biochemical techniques. Molecular methods involve examining the DNA of the bacterium, either by using a technique to map certain important characteristics of an organism's genome or by sequencing a portion of the organisms DNA. Results are then compared to a database of known bacteria, resulting in a match that allows identification. DNA sequencing has become so standard and straightforward that it is now often easier and quicker than traditional biochemical methods. The currently available molecular basis of identifications of *Vibrio* species include those of randomly amplified polymorphic DNA, 16S rDNA gene-based identification, use of oligonucleotide probes, pulse field gel electrophoresis (PFGE), and use of multiplex PCR. Several protocols (Fig. 9.1) are available for the identification of *Vibrio* species, some of them have been described in this chapter.

9.1.1. Testing of Serology

The primary basis of classification of strains of *Vibrio* species is a serotyping scheme, which depends on the antigenic properties of the somatic (O) and capsular (K) antigens. The serotyping scheme for *Vibrio* is a combination of O and K antigens and serotyping is done using commercially available antisera that include 11 different O antigens and 71 different K types. The serotyping scheme was developed using strains of clinical origin. In addition to serotyping, phage typing can be conducted using 46 phages belonging to morphological groups II, IV, and V.

9.1.2. Multilocus Sequence Analysis

Multilocus sequence analysis is a useful tool for the identification of *Vibrio* at genetic level. In this, various house keeping genes like *rpoA*, *recA*, and *pyrH* genes are targeted [3]. After amplification and sequencing of these genes, the strains can be identified and phylogeny of these strains can be compared with the former polyphasic taxonomic studies like 16S rRNA based phylogeny of *Vibrio*. The data generated by this study are well suited to be used for the rapid detection and identification of pathogenic *Vibrios* in the environment through real-time PCR. The data could be an alternative to 16S rRNA gene sequences in studies of the ecology and community dynamics of *Vibrio* in environmental as well as clinical samples. The advantages of studying the loci mentioned above is that they belong to the bacterial core genome, have a high phylogenetic signal and are single copy genes and that different species of *Vibrios* have different gene sequences that thus enable the reliable identification of these organisms.

9.1.3. Pulse Field Gel Electrophoresis

PFGE is a useful technique for the molecular identification of *Vibrio* species. PFGE banding patterns distinguished between and identified diversity within each of the major MEE types [4]. The groupings obtained by separation of strains by PFGE patterns agree with the pattern obtained by ribotyping. Although the PFGE patterns of *Vibrio cholerae* may be too numerous and analysis of these patterns may be too complex to be used in a general typing scheme, the variety that they offer is of particular value in investigations of epidemics. PFGE appears to be the most discriminating of several molecular subtyping methods used and it is reproducible, relatively stable over time, and is relatively rapid in comparison with MEE and methods requiring DNA hybridization.

9.1.4. Multiplex Polymerase Chain Reaction

Currently multiplex PCR methods have been developed to characterize the *Vibrio* species using a single PCR for many characteristics. The advantages for this PCR assay is that it overcomes the time consuming, laborious, expensive laboratory practices of biochemical basis of identification of *Vibrio* and also distinguishes among highly similar biochemical properties of *V. cholerae*, *Vibrio mimicus*, and other *Vibrio* spp., with *Aeromonas*. A septaplex PCR

protocol has been described here for the rapid identification of *Vibrio* spp. along with their virulence genes [5].

9.1.5. Analysis of Full Length *toxR* Gene

The *toxR* gene of *Vibrio* species code for a transmembrane DNA-binding protein which activates the transcription of the cholera toxin operon and a gene (*tcpA*), a major subunit of pilus colonizing factor [6]. However, the full-length sequence of *toxR* gene (i.e., 1,000 bp) from type strains of different *Vibrios* provide an insight for evaluating phylogenetic relatedness and rapid species specific identification and differential detection of unknown *Vibrios* [7]. Because of the greater sequence variation among species, the use of the *toxR* gene becomes more effective than the 16S rRNA gene in distinguishing different species of *Vibrio*.

9.2 Materials

9.2.1. Multilocus Sequence Analysis

1. DNA extraction kit
2. *rpoA*, *recA* and *pyrH* primers
3. dNTPs
4. Taq DNA polymerase
5. PCR reaction buffer
6. PCR tubes and tips

9.2.2. Pulse Field Gel Electrophoresis

1. Wash buffer (1 M NaCl, 10 mM Tris [pH 8.0], 10 mM EDTA)
2. Chromosomal grade agarose (Bio-Rad, Richmond, California)
3. Plug mold (Bio-Rad)
4. Lysis buffer (1 M NaCl, 10 mM Tris [pH 8.0], 100 mM EDTA, 0.5 % Sarkosyl, 0.2 % sodium deoxycholate, 1 mg of lysozyme per ml, 2 µg of RNase per ml)
5. ESP buffer (0.5 M EDTA, 1 % Sarkosyl, 1 mg of proteinase K per ml)
6. TE (10 mM Tris [pH 8.0], 1 mM EDTA)
7. 0.1 M phenylmethylsulfonyl fluoride
8. NotI buffer (150 mM NaCl, 10 mM Tris [pH 8.0], 10 mM MgCl₂)
9. NotI (New England Biolabs, Inc., Beverly, Mass.)
10. Fast-lane agarose gel (FMC, Rockland, Maine)
11. 0.5× TBE (10× TBE is 89 mM Tris base, 89 mM boric acid, and 2.5 mM disodium EDTA)

Table 9.1
Different primers used in multiplex PCR

Target gene	Primer sequence (5'–3')	Amplicon size (bp)	Gene accession no.	Primer site
O139 rfb-F	AGCCTCTTTATTACGGGTGG	449	Y07786	12288–12307
O139 rfb-R	GTCAAACCCGATCGTAAAGG	449	Y07786	12717–12736
O1 rfb-F	GTTTCACTGAACAGATGGG	192	X59554	13195–13213
O1 rfb-R	GGTCATCTGTAAGTACAAC	192	X59554	13368–13386
ISR rRNA VC-F	TTAAGCSTTTTCRCTGAGAATG	295	AF114723	227–248
ISR rRNA VC-R	AGTCACTTAACCATAACAACCCG	295	AF114723	501–522
ctxA F	CGGGCAGATTCTAGACCTCCTG	564	X00171	588–609
ctxA R	CGATGATCTTGGAGCATTCCCAC	564	X00171	1129–1151
toxR F	CCTTCGATCCCCTAAGCAATAC	779	M21249	277–298
toxR R	AGGGTTAGCAACGATGCGTAAG	779	M21249	1034–1055
tcpA-F Clas/El Tor	CACGATAAGAAAACCGGTCAAGAG	620	X64098	3379–3402
tcpA-R Class	TTACCAAATGCAACGCCGAATG	620	X64098	3977–3998
tcpA-R El Tor	AATCATGAGTTCAGCTTCCCGC	823	X74730	3235–3256
Sxt-F	TCGGGTATCGCCCAAGGGCA	946	AFO99172	90–109
Sxt-R	GCGAAGATCATGCATAGACC	946	AFO99172	1016–1035

12. Bacteriophage lambda DNA ladders (FMC)

13. Ethidium bromide (2 µg/ml in water)

9.2.3. Multiplex Polymerase Chain Reaction

1. Luria Bertani agar media
2. PCR buffer
3. Magnesium chloride
4. dNTPs
5. Taq polymerase
6. Autoclaved Milli Q water
7. Primers (Table 9.1)

9.2.4. Analysis of Full-Length *toxR* Gene

1. Genomic DNA extraction kit
2. PCR buffer
3. Magnesium chloride
4. dNTPs
5. Taq polymerase
6. Autoclaved MilliQ water
7. Primers (Foreward primer *toxRPV* [5'-ATGACTAATAT-CGGCAC-3'] and Reverse primer *toxS-R* [5'-GCCATTCTT-TAGAGGTCARNAVYTGyTC-3'])
8. Nucleic acid purification kit
9. BigDye® Terminator v3.1 cycle sequencing kit

9.3 Methods

9.3.1. Multilocus Sequence Analysis

PCR	
Reaction mixture	Cycling conditions
10× buffer-5.0 µl	Denaturation-94 °C for 5 min
10 mM dNTPs mixture-2 mM each	Denaturation-94 °C for 30 s
10 µM Forward Primer-2.5 µl	Annealing-55 °C for 30 s
10 µM Reverse Primer-2.5 µl	Extension-72 °C for 2 min
2.5 U/µl of Taq polymerase-1 µl	Extension-72 °C for 5 min
Water-29.5 µl	Holding-4 °C forever
Template DNA-2 µl	

} 30 cycles

Then the amplified PCR products were run in a gel to check amplification from where these products were purified using PCR Clean-up kit® Sigma–Aldrich and were amplified by chain termination technique using ABI PRISM 3100 genetic analyzer (Applied Biosystem). Then the raw sequences were processed using BIOEDIT software and the phylogenetic tree was constructed using MEGA 5.0.

9.3.2. Pulse Field Gel Electrophoresis

1. Cultures were incubated in 15 ml of heart infusion broth at 37 °C for 1–1.5 h with aeration until growth reached an optical density of 0.6 at 610 nm.
2. Cells (10 ml) were harvested by centrifugation and were-washed with 10 ml of wash buffer (1 M NaCl, 10 mM Tris [pH 8.0], 10 mM EDTA).
3. Cells were resuspended in 1 ml of wash buffer and were warmed at 37 °C for a few minutes. Bacterial suspensions were mixed with an equal volume of 1 % chromosomal grade agarose (Bio-Rad, Richmond, Calif.) and were dispensed into a plug mold (Bio-Rad).
4. Agarose plugs were allowed to solidify on ice for 10 min. Plugs were placed in clean tubes containing 3 ml of lysis buffer (1 M NaCl, 10 mM Tris [pH 8.0], 100 mM EDTA, 0.5 % Sarkosyl, 0.2 % sodium deoxycholate, 1 mg of lysozyme per ml, 2 µg of RNase per ml).
5. Bacteria were lysed in the agarose plugs for 1 h at 37 °C. The lysis buffer was removed, and the plugs were incubated overnight in 3 ml of ESP buffer (0.5 M EDTA, 1 % Sarkosyl, 1 mg of proteinase K per ml) at 50 °C.
6. The next day the plugs were rinsed briefly with deionized water. Plugs were washed twice in 2 ml of TE (10 mM Tris [pH 8.0], 1 mM EDTA) containing 30 µl of 0.1 M phenyl methylsulfonyl fluoride for 30 min each. Plugs were washed four times in 3 ml of TE for 30 min each time.
7. If the plugs were not to be used immediately, only two washes in TE were performed; this was followed by overnight incubation in 5 ml of TE at 4 °C. A small portion of the plug (2 × 7 mm) was sliced off and was incubated for 1 h in a microcentrifuge tube in 1 ml of NotI buffer (150 mM NaCl, 10 mM Tris [pH 8.0], 10 mM MgCl₂).
8. The buffer was then replaced with 125 µl of fresh buffer containing 20 U of NotI (New England Biolabs, Inc., Beverly, Mass.), and the mixture was incubated for 4 h at 37 °C.
9. Restriction fragments were separated in a 1 % fast-lane agarose gel (FMC, Rockland, Maine) in 0.5× TBE (10× TBE is 89 mM Tris base, 89 mM boric acid, and 2.5 mM disodium EDTA) by using a CHEF-DR II system (Bio-Rad).
10. Bacteriophage lambda DNA ladders (FMC) were used as molecular mass standards.
11. A model 1,000 mini chiller (Bio-Rad) was used to maintain the temperature of the buffer at 14 °C. A ramp time of 5–50 s for 20 h at 200 V was used to maximize the separation of larger fragments. For longer gels, the run time was increased to 22 h.

12. If separation of smaller fragments was necessary, a ramp time of 1–10 s for 12 h at 200 V was used. Following electrophoresis, gels were stained for 20 min with ethidium bromide (2 µg/ml in water), destained, and visualized on a UV light box.
13. *V. cholerae* O1 isolates were separated into patterns on the basis of differences in band arrangements.
14. Differences in the presence, absence, or intensity of a band among strains were given equal weights.
15. Strains that differed by one band were assigned different pattern numbers. Pattern numbers were designated solely for discussion purposes and are not meant to imply relatedness between isolates or to fulfill a typing scheme.

9.3.3. Multiplex Polymerase Chain Reaction

1. Grow *Vibrio* strains in LB broth overnight at 37 °C.
2. Boil them for 10 min in water bath and store at –20 °C till further use which will be used as template.
3. Perform Multiplex PCR using the following protocol:

PCR

Reaction mixture	Cycling conditions	
10× buffer-10 µl		
Magnesium chloride-10 µl	Denaturation-94 °C for 5 min	
10 mM dNTPs mixture-10 µl	Denaturation-94 °C for 30 s	} 30 cycles
10 µM forward primer-4.0 µl	Annealing-55 °C for 30 s	
10 µM reverse primer-4.0 µl	Extension-72 °C for 2 min	
2.5 U/µl of Taq polymerase-1 µl	Extension-72 °C for 5 min	
Water-up to 97 µl	Holding-4 °C forever	
Template DNA-3 µl		

4. Separate the PCR product in 2.5 %, w/v of agarose by electrophoresis and stain the gel in ethidium bromide and visualize the banding pattern under UV.
5. The banding pattern will be observed as the Fig. 9.2.

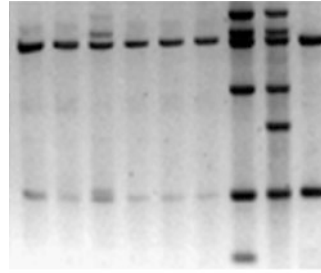


Fig. 9.2 Banding pattern of the *Vibrio* isolates subjected to septaplex PCR assay.

9.3.4. Analysis of Full-Length *toxR* Gene

1. Extract the genomic DNA of the bacterial strain from the overnight grown culture using DNeasy[®] Blood and Tissue Kit (QIAGEN GmbH, Germany) following the manufacturers protocol.
2. Amplify the *toxR* gene using the genomic DNA as template as per the following conditions:

PCR

Reaction mixture	Cycling conditions	
10× buffer-5.0 µl	Denaturation-94 °C for 2 min	
10 mM dNTPs mixture-0.8 mM each	Denaturation-94 °C for 1 min	} 30 cycles
Forward primer-0.5 µM	Annealing-51.4 °C for 1 min	
Reverse primer-0.5 µM	Extension-72 °C for 2 min	
2.5 U/µl of Taq polymerase-1 µl	Extension-72 °C for 7 min	

3. Purify the amplified *toxR* gene using the Nucleospin[®] Nucleic Acid Purification Kit (Clontech Laboratories, Inc., USA) using manufacturer's protocol.
4. Perform DNA sequencing by 1st BASE Pte. Ltd. (Singapore) using the 3730XL Genetic Analyzer together with the BigDye[®] Terminator v3.1 Cycle Sequencing Kit (both from Applied Biosystems, USA) and *toxRPV* and *toxSR* as sequencing primers.
5. Perform the *toxR* gene homology search by using the Basic Local Alignment Search Tool (BLAST) (<http://www.ncbi.nlm.nih.gov/blast>).
6. Deduce the aminoacid sequence of *toxR* using the Expert Protein Analysis System (ExPASy) Translate tool (<http://>

www.expasy.ch/tools/dna.html; Swiss Institute of Bioinformatics, Switzerland) and align with ToxR amino acid sequences from other *Vibrio*.

7. Construct the phylogenetic tree based on the sequences of *toxR* gene of other *Vibrio* species using the Molecular Evolutionary Genetics Analysis (MEGA), version 4.0 software.
8. For the construction purposes employ the neighbor-joining (NJ), p-distance method, and assess the reliability of topologies by bootstrap method with 10,000 replicates.

References

1. Ryan KJ, Ray CG (2004) Sherris medical microbiology, 4th edn. McGraw Hill, New York, NY. ISBN 0-8385-8529-9
2. Faruque SM, Nair GB (2008) *Vibrio cholerae*: genomics and molecular biology. Caister Academic Press, Norfolk, UK. ISBN 978-1-904455-33-2
3. Thompson FL, Gevers D, Thompson CC, Dawyndt P, Naser S, Hoste B, Munn CB, Swings J (2005) Phylogeny and molecular identification of *Vibrios* on the basis of multi-locus sequence analysis. Appl Environ Microbiol 71:5107–5115
4. Cameron DN, Khambaty FM, Wachsmuth IK, Tauxe RV, Barrett TJ (1994) Molecular characterization of *Vibrio cholerae* O1 strains by pulsed-field gel electrophoresis. J Clin Microbiol 32:1685–1690
5. Mantri CK, Mohapatra SS, Ramamurthy T, Ghosh R, Colwell RR, Singh DV (2006) Septaplex PCR assay for rapid identification of *Vibrio cholera* including detection of virulence and int SXT genes. FEMS Microbiol Lett 265: 208–214
6. Miller VL, Mekalanos JJ (1988) A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. J Bacteriol 170: 2575–2583
7. Orata FD, Hedreyda CT (2011) Isolation and sequence analysis of the full-length *toxR* gene of type strain *Vibrio campbellii* and use of the *toxR* gene sequence to evaluate variation and relatedness with other *Vibrio* species. Philipp Sci Lett 4:70–80

Chapter 10

Molecular Identification of Microbes: *V. methylotrophs*

Kamlesh K. Meena, Manish Kumar, and D.P. Singh

Abstract

Methylotrophic bacterial community is well known for the utilization of C1 and multiple reduced carbon substrates for their source of carbon and energy. They play an important role in the carbon cycling. The molecular methods are being described for the identification of cultivable and uncultivable methylotrophic strains. At the level of gene coding the methylotrophic function and discrimination among the strains, the tools and techniques are illustrated in details. Molecular identification of the isolates enables us to access the phylogeny and functional diversity of the methylotrophic community.

10.1 Introduction

Since microorganisms were first isolated and grown in pure culture, microbiological laboratories have needed to characterize isolates so that they can be differentiated from one another. The identification of a microbial isolate up to species level only amounts to a partial characterization of the isolate, but is still a very useful piece of information. Knowing the species allows the laboratory access to the body of knowledge that exists about that species. Schemes that can be used to describe the characteristics of a microbial isolate are essential in every branch of microbiology and their development and refinement have been constant. The advent of molecular biology in the 1980s contributed a set of powerful new tools that have helped microbiologists to detect the smallest variations within microbial species and even within individual strains. This has added an entirely new dimension to a science that was in danger of becoming constrained by its reliance on traditional laboratory techniques. In fact, the technology has progressed far beyond the level needed by most routine laboratories, where identifying the species of any isolate is likely to be sufficient. Methylotrophic bacteria are a diverse group of microbes that utilize one-carbon compounds more reduced than CO₂ as

sole energy sources and assimilate carbon at the oxidation level of formaldehyde [1, 2]. The ability to grow methylotrophically was first discovered in the early 1900s, it was not until the 1960s to 1970s that an understanding of the biochemical nature of this capability started to emerge. Bacteria able to grow on methane are a subset of the methylotrophs called methanotrophs [3]. These subpopulations of bacteria grow on a variety of carbon substrates such as methane, methanol, methylated amines, halogenated methanes, and methylated sulfur species. Methylobacteria are ubiquitous and present in diverse environments such as leaf phyllosphere [4], sea water [5], like deep-sea sediments [6], drinking water, chlorinated environments [7], hot water effluent [8], hypersaline lake [9], and fresh water lake [10]. The ability to grow on reduced C1 compounds requires the presence of unique biochemical pathway, for both energy and carbon metabolism. A key feature of aerobic methylotrophy is the role of formaldehyde as a central intermediate. In most methylotrophs, the pool of formaldehyde generated from methylotrophic substrates is split, with part being oxidized to CO₂ for energy and part being assimilated into cell carbon via one of two unique pathways, the serine cycle or the ribulose monophosphate cycle. Bacteria oxidizing CO₂ and assimilating via classical Calvin–Benson–Bassham cycle are called “Pseudomethylotrophs” or “autotrophic methylotrophs” [11]. Here we explained the molecular identification of methylotrophs; C1 carbon substrate utilizers, with tools and techniques, distinguishing different strains. Nevertheless, methods and equipment designed to help with both species identification and typing are commercially available for a range of applications.

10.1.1. Traditional Identification Method

Given the microbial nature of the disease, traditional or culture-based microbiology studies were carried out by a number of oral microbiology groups throughout the world. Over the next 2 decades or so, it was shown that the bacterial species associated with these lesions were surprisingly limited, given the number of taxa potentially able to colonize and the large number of taxa associated with periodontal lesions. This reduced diversity implies special selective pressures operating within the root canal system. While the culture-based techniques have reported 4–12 taxa per root canal when the range of taxa isolated from root canal infections as a group is taken into account, 20–30 genera are commonly isolated; of these most commonly occurring species are *Fusobacterium nucleatum*, *Streptococcus* species, *Porphyromonas* species, *Prevotella intermedia*, *Eptostreptococcus* species, Actinomyces species and *Eubacterium* species. (The genus *Eubacterium* is very broad and at present undergoing significant taxonomic revision.) The isolation and identification of these taxa lead to large numbers of studies aimed at defining which taxa were responsible for the disease, what mechanisms they used and

indeed, associating particular taxa to different aspects of root canal infections, e.g., pain, lesion size, etc. From early microscopy studies it was shown that 50 % of the oral microbiota was unculturable. Therefore, it was very possible that unculturable taxa were present in root canal infections and were potentially playing a role in the disease initiation or progression or both. These unculturable taxa fall into two broad categories. The first are taxa that need nutrients or other essential components that conventional sampling techniques, transport conditions, or laboratory media do not provide. This could be sensitivity to oxygen (i.e., very strict anaerobes) or the absolute requirement for products provided by other taxa within the root canal. This taxa is therefore broadly unknown apart from microscopy studies ; unless distinct morphology is apparent there is no way of knowing what proportion of the taxa are represented in the culture-dependent proportion of the sample. The second category contains those taxa that are known, and very often common, but for some reason cannot be cultured, i.e., they are in a dormant state and “non-culturable”. The term “viable but not culturable” (VBNC) was coined to describe this state. It is thought that cells will go into this state as a protection strategy in response to adverse environmental conditions. It is very possible that “adverse” conditions exist within root canals especially nutrient deprivation and this may be another explanation for the limited taxa isolated for individual root canal infections. While microbiologists may have suspected that a number of taxa were present and unculturable (for whatever reason) there was very little that could be done other than using complex media to mimic the conditions present at the site of isolation or indeed use co-culture strategies. At the end of the day they had to be able to culture the taxa before they could identify or indeed characterize them.

10.1.2. Molecular Identification

With the advent of “molecular biology,” microbiologists had another avenue to pursue with respect to understanding the microbiology. Shortly after Kary Mullis described a polymerase chain reaction (PCR) technique, the flood gates opened with respect to what was possible in the world of microbial detection and identification. The application of PCR and sequencing (and associated database construction and searching software) revolutionized the detection and identification of bacteria. PCR is a technique, which uses a DNA polymerase enzyme to make a huge number of copies of virtually any given piece of DNA or gene. It facilitates a short stretch of DNA (usually fewer than 3,000 bp) to be amplified by about a millionfold. In practical terms it amplifies enough specific copies to be able to carry out any number of other molecular biology applications, e.g., size determination (in bases) and its nucleotide sequence. The particular stretch of DNA to be amplified, called the target sequence, is

Table 10.1

List of primers are being used for identification of functional groups of methylotrophic bacteria

Genes	Targets	Primer sequence (5'–3')
<i>pmoA</i>	Particulate methane mono oxygenase	GGNGACTGGGACTTCTGG GAAGSCNGAGAAGAASGC [12]
<i>mmoX</i>	Methane mono oxygenase	GGCTCCAAGTTCAAGGTCAG TGGCACTCGTAGCGCTCCGGCTCG [12]
<i>fae1</i>	Formaldehyde activating enzyme	GTCGGCGACGGCAAYGARGTCG GTAGTTGWANTYCTGGATCTT [13]
<i>fae2</i>	Formaldehyde activating enzyme	GCACACATCGACCTSATCATSGG CCAGTGRATGAACGCCAC [13]
Serine	Serine-glyoxylate amino transferase of Serine pathway	ATGGC(AGCT) ATGAA(CT)AT(ACT)CC(AGCT) ATGGATA(AGCT)GG(AG)AA(AG)AA(AGCT)CC-(CT) TC(AG)TC [14]
HPS	Hexulose phosphate synthase of RuMP pathway	ATGAAGCTICAGGTC(A/G/T)GCIATC(A/T)GA- CC(A/G/T)GCGTGCATCTCC(A/G/T)ACGAA [15]
Mau	Methylamine dehydrogenase	ARK CYT GYG ABT AYT GGG G GAR AYW GTG CAR TGR TAR GTC [16]
<i>mxnF</i>	Methanol dehydrogenase	GCGGCACCAACTGGGGCTGGT GGGCAGCATGAAGGGCTCCC [9]

identified by a specific pair of DNA primers (Table 10.1), oligonucleotides usually about 20 nucleotides in length which designate the outer limits of the amplification product. Given that there are about 500 bacterial taxa present in the oral cavity, the range and complexity of the techniques utilized to identify this very diverse microbiota is bewildering. Molecular biology techniques have lead to new approaches for bacterial identification. The use of nucleotide sequence data from 16S ribosomal RNA genes (among others), now makes it possible not only to identify but also to infer phylogeny for all organisms on Earth. Phylogeny is defined as the evolutionary relationships within and between taxonomic levels, particularly the patterns of lines of descent, in a sense a family tree spanning 3.5 billion years. Therefore, within reason a single methodology can be used to identify any bacterial isolate from any environment.

16S (small subunit) rRNA gene was selected as a candidate molecule for a number of reasons, viz., (a) it is present in all organisms and performs the same function, (b) its sequence is sufficiently conserved and contains regions of conserved, variable, and hyper variable sequence, (c) it is of sufficient size (ca. 1,500 bases) to be relatively easily sequenced but large enough to

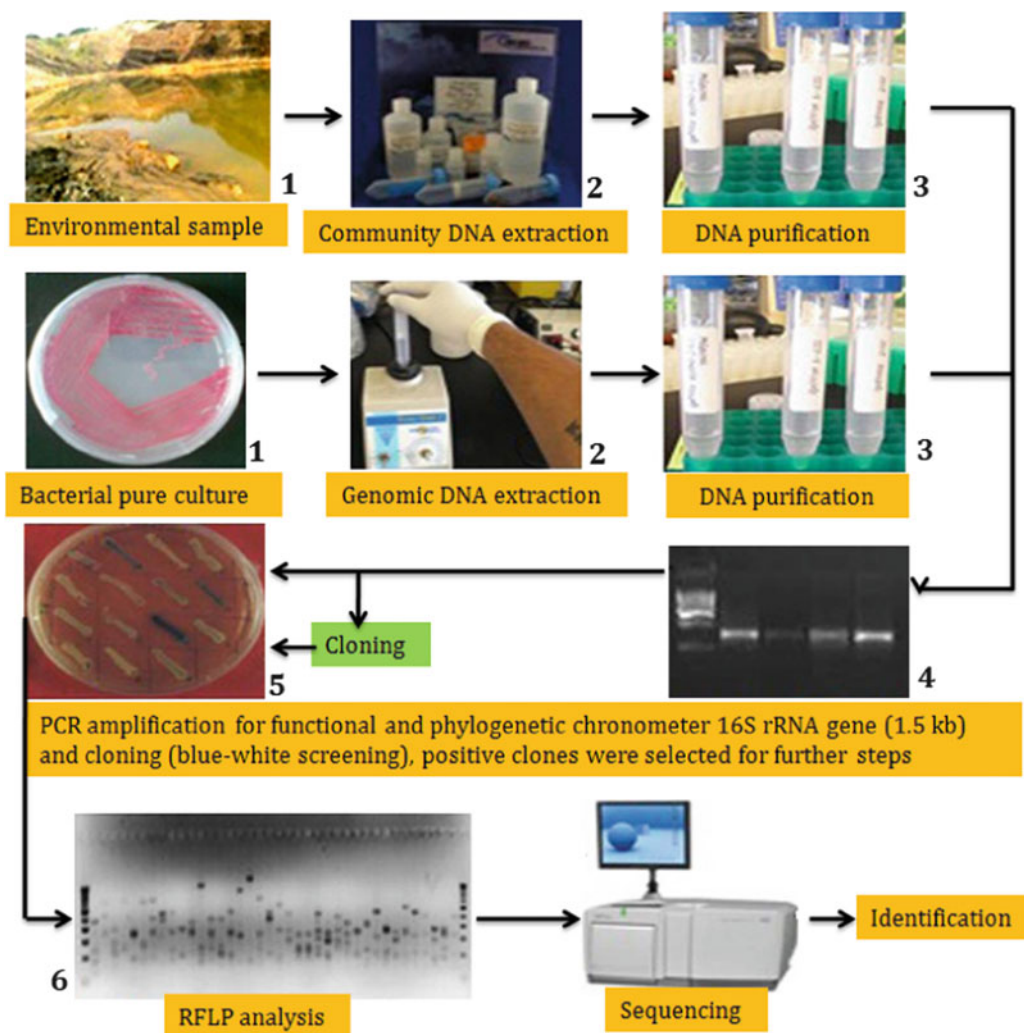


Fig. 10.1 Pictorial representation of different steps [1–3, 7, 8, 11, 12, 17] used for molecular identification of both cultivable and uncultivable methylotrophic bacteria

contain sufficient information for identification and phylogenetic analysis. Denaturing gradient gel electrophoresis, Cloning, Sequencing, Pyrosequencing, and real-time PCR gives the insight for metagenomic sequences. Figure 10.1 describes the different steps used for molecular identification of both cultivable and uncultivable methylotrophic bacteria. For the identification of uncultivable methylotrophic bacteria, the community DNA has been extracted directly from the environmental samples (soil, water, and sediments) whereas the pure culture was used for genomic DNA extraction for molecular identification of cultivable methylotrophic bacteria.

10.2 Materials

1. SET buffer (75 mM NaCl, 25 mM EDTA, and 20 mM Tris)
2. Lysozyme (10 mg ml⁻¹)
3. 0.8 % Agarose gel
4. Proteinase K and RNase
5. Ethidium bromide for staining the gel
6. UltraClean Soil DNA isolation kit
7. Primers pA 5'-AGAGTTTGATCCTGGCTCAG3' (*E. coli* position 8–27) and pH 5'-AAGGAGGTGATCCAGCCGCA3' (*E. coli* position 1525–1544)
8. 100 μM (each dNTP) dATP, dCTP, dTTP, and dGTP
9. U Taq polymerase
10. Thermal cycler
11. Primers mxoF-1003 (5'GCGGCACCAACTGGGGCTGGT3; forward) and mxoR-1561 (5'GGGCAGCATGAAAGGGCTCCC3; reverse)
12. For the environmental DNA, GC clamp (CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG G) attached at 5' end one of the mxoF or mxoR primers
13. Alpha Imager analysis system
14. Restriction endonucleases *Hae*III, *Msp*I, and *Eco*RI
15. 6.5 %, w/v acrylamide:bis acrylamide (37.5:1)
16. 1.0× TAE, pH 7.4 (0.04 M Tris-base, 0.02 M sodium-acetate, 1 mM EDTA)
17. 7 M Urea
18. 40 %, v/v formamide
19. Quiaquick PCR purification kit
20. Fluorescent terminators
21. Sterile milli-q purified water
22. 2 μl MgCl₂ (25 mM)
23. 2 μl of SYBR Green master mix (20pmol)
24. 70 % Ethanol
25. Absolute ethanol (99.9 %)
26. Water saturated phenol

10.3 Method

10.3.1. Genomic and Community DNA Extraction and PCR Amplification of 16S rRNA and *mxoF* Gene

1. Log phase cultures from NMS broth were used for genomic DNA isolation. Pelleted cells from 1.5 ml media were resuspended in 0.5 ml SET buffer (75 mM NaCl, 25 mM EDTA, and 20 mM Tris) with 10 μ l of lysozyme (10 mg ml⁻¹) and genomic DNA was extracted following the method described by Pospiech and Neumann [18].
2. The integrity and concentration of purified DNA was determined by agarose gel electrophoresis on 0.8 % agarose gel stained with ethidium bromide.
3. DNA from sediment samples were extracted using UltraClean Soil DNA isolation kit. The concentration of genomic DNA was adjusted to a final concentration of 50 ng μ l⁻¹ for PCR amplification. 16S rRNA gene was partially amplified using primers pA 5'-AGAGTTTGATCCTGGCTCAG3' (*E. coli* position 8–27) and pH 5'-AAGGAGGTGATCCAGCC-GCA3' (*E. coli* position 1,525–1,544) [17].
4. The amplification was carried out in a 100- μ l volume by mixing 50–90 ng template DNA with the polymerase reaction buffer (10 \times); 100 μ M (each) dATP, dCTP, dTTP, and dGTP; primers pA and pH (20 ng each), and 1.0 U Taq polymerase using following conditions: initial denaturation at 94 °C for 1.5 min, 35 cycles consisting of 95 °C for 1 min, 55 °C for 1 min, 72 °C for 1 min, and final extension 72 °C for 5 min on a thermal cycler.
5. The *mxoF* gene was used to identify or to authenticate populations capable of methanol oxidation downstream of formaldehyde. The *mxoF* gene in the isolates was partially amplified using specific primers, *mxoF*-1003 (5'GCGGCACCAACT-GGGGCTGGT3; forward) and *mxoR*-1561 (5'GGGCAGC-ATGAAAGGGCTCCC3; reverse). For the environmental DNA, GC clamp (CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG G) attached at 5' end with one of the primer, was used for increasing the separation of DNA bands in DGGE gel.
6. The PCR condition was similar to 16S rDNA amplification. PCR products were separated on 1.5 % agarose gel stained with ethidium bromide and documented in Alpha Imager analysis system.

10.3.2. Amplified Ribosomal DNA Restriction Analysis (ARDRA) and RFLP

1. Following 16S rDNA amplification, the products were digested with selected restriction enzymes having different restriction sites. Approximately 1 µg of PCR-amplified 16S rDNA fragments were restricted with three different endonucleases *Hae*III, *Msp*I, and *Eco*RI separately, incubated at 37 °C for overnight, and resolved on 2 % agarose gels.
2. The *mx*aF gene PCR products were digested with *Hae*III and *Eco*RI and separated through PAGE. The banding pattern obtained were visualized by ethidium bromide staining and documented in Alpha Imager documentation analysis system.
3. Different phylotypes or operational taxonomic units were obtained by similarity and clustering analysis using the software, NTSYSpc-2.02e. Similarity among the isolates was calculated by Jaccard's coefficient [11], and dendrogram was constructed using UPGMA method.

10.3.3. Cloning (Electroporation Method)

Cloning is an important step in the molecular identification of any organisms. Basically cloning is a three steps process, i.e., competent cell preparation, ligation, and transformation. Competent cells can be prepared by two different methods like (electroporation and chemical method). Both the processes are being described here under below.

10.3.3.1. Preparation of Electro-competent DH10B Cells

1. Pick a colony of DH10B and inoculate into a 5 ml of LB broth. Grow overnight at 37 °C with shaking.
2. Next morning, add 1 % of the O/N culture into 500 ml of LB medium and grow at 37 °C with shaking until OD 600 reaches ~ 0.7 (this takes ~ 2 h).
3. Cool cells in cold room on ice for ~ 20 min.
4. Spin centrifuge rotor for 5 min to precool to 4 °C.
5. Pour cells into 2 precooled GS3 bottles, 250 ml each. Spin at 5,000 rpm for 15 min at 4 °C.
6. Decant supernatant and resuspend cells in 500 ml of sterile ice-cold water.
7. Spin at 5,000 rpm for 15 min at 4 °C. Decant supernatant and resuspend cells in 250 ml of sterile ice-cold water.
8. Centrifuge at 5,000 rpm for 15 min at 4 °C. Decant supernatant and resuspend cells in 125 ml of sterile ice-cold water.
9. Spin at 5,000 rpm for 15 min at 4 °C. Decant and resuspend cells in 10 ml ice-cold 10 % glycerol. Transfer cells into a 50 ml centrifuge tube and centrifuge on table-top centrifuge, at max speed for 15 min at 4 °C. Decant supernatant and resuspend cells in 0.5–1 ml sterile 10 % glycerol.

Table 10.2
Ligation reaction mixture

Reagents	Quantity (μl)
Dephosphorylated vector	2.0
5× Ligation buffer	2.5
3–8 kb genomic DNA fragments	12.5
Deionized water	11.0
T4 DNA ligase	2.0
Total volume	30.0

10. Aliquot 100 μl of cells into each pre-cooled microcentrifuge tubes on ice and store electro-competent cells at -70°C until use.

10.3.3.2. Ligation

1. Set up the ligation reaction as follows (Table 10.2) and incubate the ligation mixture at 22°C overnight.
2. Inactivate the ligation by heating at 70°C for 10 min and use the ligation mixture for the electrotransformation of *E. coli* DH10B cells.

10.3.3.3. Transformation

1. Prior to electroporation, ligation mixture must be precipitated with ethanol or diluted to prevent the samples from causing an arc to jump across the cuvette upon application of the pulse.
2. Thaw an aliquot of DH10B cells on ice. When cells are thawed, add 1–10 μl of ligation mixture to the cells and mix by tapping gently.
3. Carefully pipette the cell/DNA mixture into a chilled 0.1 cm cuvette. Gently tap the cuvette to ensure that the cell/DNA mixture makes contact all the way across the bottom of the cuvette chamber. Avoid formation of bubbles.
4. Wipe the outside of the cuvette with a tissue to dry it, place it in the electroporation chamber, and apply pulse. For BioRad GenePulser[®] II electroporator, the recommended pulse conditions are 2.0 kV, 200 Ω, and 25 μF.
5. Immediately after pulsing, add 900 μl of SOC medium and transfer the solution to a microcentrifuge tube. Delaying this transfer can seriously reduce the survival of transformed cells.
6. Incubate at 225 rpm (37°C) for 1 h with shaking.

7. Spread the cells on LB agar plates containing ampicillin ($100 \mu\text{g ml}^{-1}$), X-gal ($20 \mu\text{g ml}^{-1}$), and IPTG ($40 \mu\text{g ml}^{-1}$).
8. Incubate plates overnight at 37°C , to permit the color to develop sufficiently to distinguish blue colonies from white.

10.3.4. DGGE (Denaturing Gradient Gel Electrophoresis) Profiling

1. The *mxhF* PCR products were purified using SV-PCR purification kit according to the manufacturer's instructions. DGGE was performed as described by Muyzer et al. [19], briefly, PCR products were separated on a 1.0-mm thick, vertical polyacrylamide gel (6.5 %, w/v acrylamide: bisacrylamide (37.5:1); Bio-Rad) prepared with and electrophoresed in 1.0TAE, pH 7.4 (0.04 M Tris-base, 0.02 M sodium-acetate, 1 mM EDTA) at 60°C and constant voltage of 150 V for 16 h.
2. A denaturing gradient of 100 % denaturant corresponded to 7 M urea plus 40 %, v/v formamide. The gels were loaded with 30 μl of PCR product, depending on the band intensity following electrophoresis on 1.5 % agarose gels.
3. DGGE Gels were stained for 20 min in water containing $0.5 \mu\text{g ml}^{-1}$ ethidium bromide. Images were recorded in system. DNA bands migrating to the same position in the gel were assumed to be identical amplicons.

10.3.5. 16S rRNA and *mxhF* Gene Sequencing

1. The PCR amplified 16S rDNA products were purified with a Quiaquick purification kit . The DNA sequence was double checked by sequencing both strands using primers pA and pH for forward and reverse reaction, respectively.
2. The nucleotide sequences were dideoxy cycle sequenced with fluorescent terminators and run in 3130xl Applied Biosystems ABI prism automated DNA sequencer.
3. The DGGE bands were excised from the gel using a sterile scalpel and incubated in 60 μl sterile milli-q purified water for 24 h at 4°C . After this period, the DNA has diffused out of the gel and the solution can be used as the template in a re-amplification PCR.
4. Re-amplification was performed using the original primers but modified PCR programs and run on DGGE to confirm its identity. Only pure bands were used for the sequencing by amplifying with primers without a GC clamp.
5. PCR products for sequencing were purified and sequenced using ABI prism sequencer. The representatives of the den-drogram constructed from *mxhF*-RFLP (PAGE) pattern were also sequenced with fluorescent terminators and run in same DNA sequencer.

6. Sequences were aligned to related sequences in the public databases: NCBI. Phylogenetic tree were calculated and drawn using Neighbor-joining algorithm in the MEGA 4.0 software. For the tree construction, the aligned positions of mxaF gene sequences (both from cultures and from the DGGE bands) were compared with sequences from the database.

10.3.6. BLAST Search and Phylogenetic Analysis

1. The partial 16S rDNA sequences, mxaF gene sequences of isolated strains, and mxaF gene sequences from environmental DNA were compared with those available in the databases.
2. Identification was based on sequence similarity of $\geq 97\%$ with that of public database sequences, NCBI by BLAST homology. Sequence alignment and comparison was performed using the multiple sequence alignment program CLUSTALW2 with default parameters and the data converted to PHYLIP format. Minor modifications were done manually on the basis of conserved domains and columns containing more than 50 % gaps were removed. The phylogenetic tree was constructed on the aligned datasets using neighbor joining (NJ) method using the program MEGA 4.0.2. Bootstrap analysis was performed as described by on 1,000 random samples taken from the multiple alignments.

10.3.7. Quantification by Real Time PCR

1. Quantification of mxaF gene was done in terms of copies using LightCycler software 3.5 based on “second derivative maximum method” in which exponential phase of amplification curve is linearly related to a starting concentration of template DNA molecules.
2. Quantitative PCR using SYBR GreenI technology with the primers mxaF and mxaR was carried out amplifying five environmental DNA from sediment, negative control, and five plasmid DNA standards. Mastermix was prepared as 14 μl of sterile water, 2 μl MgCl_2 (25 mM), 1 μl of each primer (20 pmol), 2 μl of SYBR Green master mix (20 pmol) [Roche diagnostics], and 50 ng of gDNA. Amplification program applied :10 min of denaturation at 95 °C, followed by 40 cycles of four-segment amplification were accomplished with: 15 s at 95 °C for denaturation, 10 s at 55 °C for annealing, 20 s at 72 °C for elongation, and 5 s at 83 °C appended for a single fluorescence measurement above melting temperature of possible primer dimers. This fourth segment eliminates a nonspecific fluorescence signal and ensures accurate quantification of desired product. Subsequently, a melting step was performed consisting of 10 s at 95 °C, 10 s at 60 °C, and slow heating with a rate of 0.1 °C per second up to 99 °C with continuous fluorescence measurement.

References

1. Anthony C (1982) The biochemistry of methylotrophs. Academic Press, Inc., London
2. Bratina B, Hanson RS (1993) Methylo-trophy. In: Lederberg J (ed) Encyclopedia of microbiology. Academic Press, Inc., San Diego, California, pp 121–127
3. Brusseau GA, Buliygina ES, Hanson RS (1994) Phylogenetic analysis and development of probes for differentiating methylotrophic bacteria. Appl Environ Microbiol 60:626–636
4. Olivier N, Noyes E, Kalyuzhnaya MG, Lidstrom ME, Chistoserdova L (2005) Bacterial populations active in metabolism of C₁ compounds in the sediment of Lake Washington, a freshwater lake. Appl Environ Microbiol 71(11):6885–6899
5. Wang X, Sahr F, Xue T, Sun B (2007) *Methylobacterium salsuginis* sp. nov., isolated from sea water. Int J Syst Evol Microbiol 57: 1699–1703
6. Wang P, Wang F, Xu M, Xiao X (2004) Molecular phylogeny of methylotrophs in a deep-sea sediment from a tropical west Pacific Warm Pool. FEMS Microbiol Ecol 47:77–84
7. Hiraishi A, Shimada K (2001) Aerobic anoxygenic photosynthetic bacteria with zinc-bacteriochlorophyll. J Gen Appl Microbiol 47:161–180
8. Bodrossy L, Murrell JC, Dalton H, Kalman M, Puskas LG, Kovacs KL (1995) Heat-tolerant methanotrophic bacteria from the hot water effluent of a natural gas field. Appl Environ Microbiol 10:3549–3555
9. McDonald IR, Murrell JC (1997) The methanol dehydrogenase structural gene *mdhA* and its use as a functional gene probe for methanotrophs and methylotrophs. Appl Environ Microbiol 63:3218–3224
10. Muyzer G, De Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl Environ Microbiol 59:695–700
11. Jaccard P (1912) The distribution of the flora in the alpine zone. New Phytol 11:37–50
12. Holmes AJ, Owens NJP, Murrell JC (1995) Detection of novel marine methanotrophs using phylogenetic and functional gene probes after methane enrichment. Microbiology 141:1947–1955
13. Kalyuzhnaya MG, Lidstrom ME, Chistoserdova L (2004) Utility of environmental probes targeting ancient enzymes: methylotroph detection in Lake Washington. Microb Ecol 48:436–472
14. Tairo HI, Toyokdzu Y, Yoshikazu I, Toshio M (1996) Cloning and expression of the gene for serine-glyoxylate aminotransferase from an obligate methylotroph *Hyphomicrobium methylovorum* GM2. Eur J Biochem 24:1–5
15. Mitsui R, Sakai Y, Yasueda H, Kato N (2000) A novel operon encoding formaldehyde fixation: the ribulose monophosphate pathway in the gram-positive facultative methylotrophic bacterium *Mycobacterium gastri* MB19. J Bacteriol 182:944–948
16. Neufeld JD, Schafer H, Cox MJ, Boden R, McDonald IR, Murrell JC (2007) Stable-isotope probing implicates *Methylophaga* spp. and novel Gammaproteobacteria in marine methanol and methylamine metabolism. ISME J 1:480–491
17. Edwards U, Rogall T, Blocker H, Emde M, Bottger EC (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. Nucleic Acids Res 17: 7843–7853
18. Pospiech A, Neumann B (1995) A versatile quick-prep of genomic DNA from gram positive bacteria. Trends Genet 11:217–218
19. Lidstrom ME (2006) Aerobic methylotrophic prokaryotes. Prokaryotes 2:618–634

Chapter 11

Preservation and Maintenance of Microbial Cultures

Sudheer Kumar, Prem L. Kashyap, Ruchi Singh, and Alok K. Srivastava

Abstract

The preservation and maintenance of microbial cultures require special and careful attention, reliable preservation and appropriate quality control to ensure that recovered cultures perform in the same way as the original cultures. This requires a high degree of expertise in the maintenance and management of microbial cultures at ultralow temperatures, or as freeze-dried material, to secure their long-term integrity and relevance for future research, development, and conservation. This chapter outlines some of the important procedures and protocols involved in the conservation, preservation, and maintenance of microbial cultures.

11.1 Introduction

The majority of the biomass and biodiversity of life on the Earth is possessed by microbes. They play pivotal roles in biogeochemical cycles and harbor novel metabolites and genes that have industrial and agricultural uses [1, 2]. For these reasons, the conservation and preservation of microbes is the most important and high priority task for most microbiological laboratories engaged in research, teaching, or industrial application. Secondly, the availability of quality and authentic microbial cultures are a significant advantage for promoting research, standardization, efficiency, and laboratory safety [3]. Therefore, it is important that microbial resources should be preserved in a physiologically and genetically stable state.

The preservation of microorganisms by different methodologies has been employed for decades [4]. The primary methods of culture preservation are continuous growth, drying, and freezing. Continuous growth methods, in which cultures are grown on agar, typically are used for short-term storage. Such cultures are stored at temperatures from 5 to 20 °C to increase the interval between subculturing. The methods are simple and inexpensive

because no specialized equipment is required. Drying is the most useful method of preservation for cultures that produce spores or other resting structures. Silica gel, glass beads, and soil are substrata commonly used in drying. Fungi have been stored successfully on silica gel for up to 11 years [5]. Drying methods are technically simple and also do not require expensive equipments. Freezing methods, including cryopreservation are versatile and widely applicable. Most of microorganisms can be preserved, with cryoprotectants, in liquid nitrogen vapor or in standard ultra-low temperature freezers [6]. With freeze-drying or lyophilization, the microbial cultures are frozen and subsequently dried under vacuum. The method is highly successful with cultures that produce mitospores. Freeze-drying and freezing below -135°C are excellent methods for permanent preservation of bacteria, actinomycetes, yeasts, and fungi [7]. For long-term storage, cultures are usually preserved by lyophilization or by ultra-freezing [3, 8, 9]. In these methods, two basic approaches are employed to slow down the rate of deleterious reactions in microbial culture. The first is to lower the temperature which decreases the rate of all chemical reactions. This can be done using refrigerators and liquid nitrogen freezers. The second option is to remove water from the culture, a process which can be tricky and involves sublimation of water using a lyophilizer [10]. Briefly, in this chapter, we provide an overview of some of the important procedure involved in the short-/long-term maintenance and conservation of microbial cultures.

11.2 Periodical Transfer to Fresh Media

Microbial strains can be maintained by periodically preparing a fresh culture from the previous stock culture. The culture medium, the storage temperature, and the time interval at which the transfers are made vary with the species and must be ascertained beforehand. The temperature and the type of medium chosen should support a slow rather than a rapid rate of growth, so that the time interval between transfers can be as long as possible. Many of the microbes remain viable for several weeks or months on a medium like nutrient agar or potato dextrose agar (PDA). This method is generally used for maintenance of cyanobacteria. The periodic transfer method has the disadvantage of failing to prevent changes in the characteristics of a strain due to the development of variants and mutants.

11.3 Refrigeration

Microbes can be preserved for a short period of time at 4 °C. The cultures require frequently in active growth on agar slants or plates can be stored in a refrigerator and precaution has to be taken to avoid contamination. Cultures should be prepared using standard techniques and then sealed before storing. For slants, it is recommended to use screw-capped tubes. For cultures on Petri dishes, the plates need to be sealed with Parafilm. Sealing the plates not only helps to prevent molds from sneaking into the plates, but it also slows the drying of agar. For short term over a week or two, cultures can be stored as stabs in small, flat-bottomed screw-capped vials. In this technique, vials are filled with a small amount of agar medium (e.g., 1 ml) and sterilized. Microbes (e.g., bacteria) are then introduced into the solidified agar with a sterile needle. The culture is incubated overnight and then stored at 4 °C. Cultures stored in stabs are more resistant to drying and contamination, but they will lose viability more quickly than frozen stocks. The length of time a stab can remain viable is dependent upon the strain.

11.4 Mineral Oil or Liquid Paraffin Storage

Covering the fresh growth in media slants with sterile mineral oil or liquid paraffin can preserve many bacteria and fungi. It was applied for the first time in 1914 by Limier to keep the gonorrhea agent (*Neisseria gonorrhoeae*). The method's basic idea is covering the well-grown culture on agar nutrient medium with sterile mineral oil. The most common used oil is paraffin or vaseline with layer thickness of 1–2 cm. The aim is to limit the oxygen access that reduces the microorganism's metabolism and growth, as well as to restrict the cell drying during preservation. The cell viability in this method is high as compared to frequent transfer and storage at low temperature. The preservation period for bacteria from the genera *Azotobacter* and *Mycobacterium* is from 7 to 10 years, for *Bacillus* it is 8–12 years.

11.4.1. Materials

1. High quality liquid paraffin with 0.830–0.890 specific gravity (Tyndalized at 121 °C for 20 min).
2. Sterile semisolid growth medium in universal bottles (30 ml) kept at 30° slope.
3. Metal segmented trays (37.5 × 17.5 cm² having 25 × 25 mm² marking).
4. Inoculating needle or loop.

11.4.2. Methods

1. Inoculate at least two universal bottles for each strain to be maintained.
2. Label one culture as reserve stock and the other(s) as working stock.
3. Incubate at optimum growth temperature until the organism has reached maturity.
4. Add sterile liquid paraffin (8–10 ml) to cover the slope to a maximum depth of 10 mm from top of the slant edge. While sterilization of liquid paraffin, it should not get moisture, if it absorbs the moisture it becomes turbid and whitish which leads to the deterioration of cultures during storage. Remove the moisture by gentle heating in oven so it becomes transparent.
5. Store the oiled cultures, with the screw caps loose, in metal divided racks at 15–25 °C.
6. Recovery:
 - (a) Remove a portion of the working stock culture using a sterile needle or loop.
 - (b) Drain as much oil as possible from the inoculum.
 - (c) Inoculate on fresh growth medium. It is often best to inoculate a slope so that the adhering oil can drain and the organism can grow up the slope away from the oil at the point of inoculation.
 - (d) The reserve stock culture is used only when re-preservation becomes necessary when all the inoculum has been removed, when it is contaminated, or when the shelf life for the organism has been reached.

11.5 Freezing

Freezing is a good way to store bacteria and most of fungi. Generally, the colder the storage temperature, longer the culture will remain viable. Freezers can be split into three categories: laboratory, ultralow, and cryogenic. However, ice crystal formation is the major problem, when bacteria are stored at low temperature. Ice can damage cells by dehydration caused by localized increases in salt concentration. As water is converted to ice, solutes accumulate in the residual free water and this high concentration of solutes can denature biomolecules. To lessen the negative effects of freezing, glycerol is often used as a cryoprotectant. With bacteria, adding glycerol to final concentration of 15 % will help to keep cells viable under all freezing conditions.

11.5.1. Freezing Bacteria Using Glycerol

Bacteria can be frozen using 15 % glycerol. The process is simple and requires screw cap microfuge tubes and sterile glycerol. The glycerol is diluted to 30 % and an equal amount of glycerol and culture broth are mixed, dispensed into tubes, and then frozen.

11.5.1.1. Materials

1. Centrifuge tube (2 ml)
2. Sterile cryovial
3. Cryotags
4. Autoclaved distilled water
5. Vortex mixer
6. Freezer
7. Sterile pipette tips
8. Glycerol (30 %, v/v)

11.5.1.2. Method

1. Prepare a solution of 30 % glycerol (v/v) by mixing 30 ml of glycerol with 70 ml of water. Transfer the solution to a screw cap glass bottle and sterilize by autoclaving at 121 °C for 15 min.
2. Aliquot 500 µl of sterile 30 % glycerol into sterile 2 ml sterile microfuge tubes (see *Note 1*).
3. Add 500 µl of bacterial culture to the tube and mix with the glycerol using a vortex mixer.
4. Label the tube with the organism name, strain, date, etc.
5. Place the tube in the freezer and record its location.
6. To revive stored bacteria, only a small volume of culture needs to be removed from the tube. The tube doesn't need to be thawed. Open the tube and scrape the frozen culture with a sterile pipette tip. Replace the tube into the freezer immediately. Transfer the small volume of frozen/thawed cells to an agar plate and streak. If the culture thaws, do not re-freeze it as cells are typically very sensitive to freezing and thawing. Discard the thawed culture appropriately.

11.5.2. Freeze-Drying (Lyophilization)

Freeze-drying (lyophilization) is a well-established method for long-term storage. It is a method of removing water, which not only serves as the medium for enzymatic reactions but also spontaneous negative reactions such as free radical formation. Many bacteria and spore forming fungi can be preserved very effectively by freeze-drying. By freezing the cells in a medium that contains a lyoprotectant (usually sucrose) and then pulling the water out using a vacuum (sublimation), cells can be effectively preserved. This method is laborious and requires specialized equipment, but it has the advantage of generating stock cultures that are unaffected by power outages and empty liquid nitrogen tanks.

Furthermore, if cultures are routinely shipped to other labs, freeze-dried cultures do not require special handling. The downside on freeze-drying is that not all cultures react the same way thus some experimentation is required to optimize the process for each strain.

There are four significant considerations for freeze-drying microorganisms. Culturing and preparing the cells is the first consideration. Generally this is not different than methods for typically culturing bacteria. The second aspect involves suspending the bacteria in a suitable freeze-drying medium, commonly, skim milk or sucrose is used. The third consideration is the freeze-drying process, which is extremely dependent upon the type of freeze-dryer used and the quantity of samples to be preserved. The final aspect deals with post-lyophilization storage. This process can be used to preserve bacteria, fungi, yeasts, proteins, nucleic acids, and any other molecules which may be degraded due to the presence of water.

11.5.2.1. Selection of a Freeze-Drying Medium

Preserving bacteria by lyophilization requires that the bacteria should be suspended in a medium that helps to maintain their viability through freezing, water removal, and subsequent storage. Common ingredients for this include mannitol, skim milk, and bovine serum albumin (BSA). A second component of a good medium is a lyoprotectant (sucrose and trehalose) which helps to preserve the structure of biomolecules throughout the lyophilization process (see Note 2). Generally, skim milk (20 %) and sucrose (5–10 %) are used as basic freeze-drying solutions.

11.5.2.2. Cultivation of the Bacteria

1. Freeze-drying is best performed on actively growing cells which are collected and suspended in freeze-drying medium. Cells are usually cultured in liquid medium and then collected by centrifugation. Alternatively, cells can be washed off a recently streaked agar plate. In either case, it is best to suspend and freeze-dry at high cell densities ($\sim 10^9$ /ml). Using higher cell concentrations ensures that the culture will retain at least some viable cells after prolonged storage. It is also possible to simply inoculate cultures into a freeze-drying medium and lyophilize at very low cell densities; however, the long-term survival of such preparations should be carefully scrutinized.
2. Freeze-drying of bacteria should be done in glass vials or ampoules. Plastic should never be used as water can actually diffuse across many plastics over time. The type of glass vessel used may be dependent upon the configuration of the freeze-dryer. A basic freeze-drying apparatus may simply be a high efficiency vacuum pump connected to a cold moisture trap which in turn is attached directly to the sample. In such cases, long neck, heat sealable ampoules should be used. Once

ampoules are filled, sterile cotton or glass wool is inserted into the neck of the ampoule to prevent contamination of the sample. Dried bacteria are sealed under vacuum in the ampoules with a propane or acetylene flame. Flame sealing is labor intensive, but the most secure method of preserving the samples. If the freeze-dryer has a drying chamber, such as with shelf dryers, then samples can be lyophilized in serum vials and sealed with rubber stoppers (called bungs). These stoppers often have a notched, or split, base that is inserted in the vial opening which allows water to escape during drying. Shelf freeze-dryers are often equipped with a stoppering mechanism that pushes stoppers into the vials effectively sealing the vial while under vacuum.

3. Whether ampoules or vials are used, these should be filled to not more than 1/3 volume. Smaller volumes are permissible and speed the freeze-drying process, thus dispensing 250 μ l of cell suspension into a 3-ml vial creates a high surface area as compared to the volume, which will allow for faster freeze-drying process. With vials, bungs are inserted and the samples are ready for processing.

11.5.2.3. Freeze-Drying Process

A basic freeze-drying process can be divided into three stages: freezing, primary drying, and secondary drying.

11.5.2.3.1. Freezing

The freezing of microorganism can be done by putting a prepared ampoule into a freezer, dry ice/ethanol bath, or liquid nitrogen. Rapid freezing works well for preserving cell viability; however, it makes the removal of water more difficult. When the frozen culture is placed under vacuum, the water jumps from the ice and obviously the surface of the culture loses water first followed by water in the center of the sample. For water to sublime from the interior of the sample, small pores or channels must form so it can escape. Rapid freezing tends to create a solid block where channel formation is minimum. Consequently rapidly frozen samples require greater drying times. Thus samples can be frozen more slowly by placing a rack of vials/ampoules in a ultralow freezer and allowing the culture to cool more slowly. Shelf dryers often have programmable temperature control that can be used to freeze cultures slowly as well. A slower rate of cooling results in larger ice crystal formation in the sample, which essentially creates the channels for water escape. Though different strains of bacteria may behave differently, dropping the temperature of prepared cells from ambient to -40°C over 30–60 min will typically be effective. However, before committing to freeze-drying large numbers of samples, test the freezing step first.

11.5.2.3.2. Primary Drying

Once the bacterial samples are frozen, the vacuum can be applied. Only high efficiency vacuums, i.e., pumps that can reduce the pressure to under 200 mtorr, will freeze-dry samples effectively. The key to primary drying is to raise the temperature of the sample so it is higher than the temperature of the cold trap. In basic systems, the cold trap is often a flask which is immersed in a dry ice ethanol bath. Ampoules connected to basic systems are initially cold (-70°C which is the temperature of the dry ice batch) but warm as they absorb ambient heat. The heat creates sufficient molecular motion to allow water molecules to sublime, i.e., go from solid ice to gas, as long as vacuum is present. With high efficiency vacuums, the trick is to remove water faster than the sample absorbs heat. The sublimation of the water thus keeps the bacterial solution frozen. If the sample increases in temperature too rapidly, the solution will melt which negates the value of freeze-drying.

Shelf freeze-dryers have a refrigerated condenser which serves as a cold trap. The condenser is also used to control the temperature of the shelf. For primary drying, the shelf temperature is raised so that the water in the sample sublimates, but melting does not occur. In this arrangement, a condenser may hold a temperature at -50°C while the shelf temperature is raised to -10°C . It is important that water moves from the warmer location (sample) to the colder location (condenser) without the sample melting. The use of matrix forming agents, such as bovine serum albumin (BSA) or mannitol, is very useful for helping to form a frozen sample that maintains its shape as water is removed. Without the additives, the sample would collapse. Additionally, the use of small volumes also benefits this primary drying stage in that water is more rapidly removed from small samples with large surfaces areas, such as with 0.5 ml in a 3-ml vial. Primary drying can take anywhere from 3 to 4 h for a small sample to overnight for a fully loaded shelf freeze-dryer.

11.5.2.3.3. Secondary Drying

Freeze-drying with a basic system does not allow separation between primary and secondary drying. As the frozen water is driven off from the sample, its temperature will rise to match that of ambient. Therefore, secondary drying is employed to force out residual water by increasing the temperature of the sample. In shelf dryers, the samples can be increased to 20°C for several hours prior to stoppering. It is important not to over dry the bacteria as this can be detrimental. The use of higher temperatures is also not recommended for the same reason. Following secondary drying, both vials and ampoules must be sealed. For shelf dryers with a stoppering mechanism, press the stoppers into the vials while under full vacuum. With ampoules, an acetylene or propane torch is used to heat the long neck of the ampoule to seal it. The vacuum will help to pull the glass closed.

11.5.2.3.4. Post-lyophilization Storage

Freeze-dried proteins can be stored at relatively warm temperatures as long as no moisture gets to the sample. This is not true for bacteria. Holding bacteria at temperatures above 4 °C for prolonged periods of time will dramatically decrease the viability of the cells. Bacteria which would otherwise be stable for years if kept in a refrigerator can die within a week at room temperature. Consequently, accelerated shelf life studies where the sample is held at 37 °C to mimic long-term storage conditions will not work with lyophilized bacteria.

For long-term storage, keep vials and ampoules at 4 °C. Periodically remove a vial/ampoule and assess the number of viable cells remaining. Rates of decay should be measured which will help to determine when the sample needs to be resuscitate and subsequently freeze-dried. For the most part, freeze-dried bacteria should be viable for several years.

11.5.2.4. Materials

1. Shelf freeze-drier, with T-matic shelf temperature control.
2. Preconstricted long-necked vials (2 ml) with butyl rubber bungs (heat sterilize vials at 180 °C for 2–3 h; bungs autoclaved at 121 °C for 15 min) and labeled with the strain number of the organism to be freeze-dried, batch number, and date of freeze-drying.
3. Pasteur pipettes.
4. Sterile nonabsorbent cotton plugs.
5. Air/gas glass blowers torch.
6. Glass cutter in support handle.
7. Sterile distilled water.
8. Lyophilization medium:
 - (a) 10 % skim milk (10 g dry skim milk + 100 ml deionized water) (see Note 3).
 - (b) 10 % sucrose (10 g of sucrose in 100 ml deionized water) (see Note 4).
 - (c) Lyophilization vials/tubes (see Note 5).

11.5.2.5. Method

11.5.2.5.5. Freeze-Drying Using Shelf Lyophilizer

1. Turn on the lyophilizer and start the condenser. If there is an external condenser using a dry ice/ethanol mixture then prepare this as well. The shelf can be set to 4 °C.
2. Center the vials on the shelf. This placement is important so that the stoppering plate can evenly press on the stoppers following freeze-drying.
3. Freeze the samples down to –40 °C either manually or through programmed controls. This step should take

approximately 30–60 min and varies with respect to instrument. If the rate of freezing can be controlled, then a drop of 1 °C/min is a practical rate. Once the samples reach temperature, they should be visibly frozen (clear liquid turn opaque and skim milk appears solid).

4. Allow the sample to sit at $-40\text{ }^{\circ}\text{C}$ for 1 h to ensure complete freezing. Vials at the center of a cluster may freeze more slowly than those on the outside.
5. Turn on the vacuum pump. Within 10–20 min, the vacuum should be under 200 millitorr (mtorr).
6. Once the vacuum is below 200 mtorr, increase the temperature of the shelf for primary drying, the phase associated with water sublimation. The temperature of the shelf is dependent upon the lyophilization medium. For sucrose, keep the shelf temperature at $-25\text{ }^{\circ}\text{C}$. In any case, the greater the difference in temperature between the shelf and the condenser/ice trap, the more efficient the primary drying process will be.
7. If melting of the samples occurs, then it might be necessary to empirically determine a shelf temperature. A practical means to do this involves placing a sample of the lyoprotective medium on a shelf temperature and incrementally lowering the temperature every 15 min. At some point the sample freezes. Under vacuum with a cold trap, your sample will be safe and will remain frozen. This is a practical method and is certainly not necessarily the most efficient primary drying temperature, but it should work well enough.
8. Primary drying is the longest phase of the freeze drying process. The idea is to keep the sample colder than condenser (or ice trap) but still sufficiently warm so that water sublimates rapidly. The temperature of the shelf can be raised to above the melting temperature as long as the sublimation process removes the heat flowing into the sample sufficiently fast to prevent melting and sample collapse (where the matrix literally caves in). The time for primary drying will also depend upon the volume of the sample. For bacteria, 0.25–0.5 ml sample is required. A limited number of samples (10–20) in a shelf dryer can be completed in just a couple of hours. A fully loaded dryer with several hundred samples may take longer. Safely, a primary drying period which is overnight should work, but test this first before you attempt to freeze dry large numbers of vials. As a standard guide, freeze dry overnight.
9. Samples still contain moisture following primary drying. The amount is debatable, but it somewhere between 2 % and 4 %. This moisture level needs to be reduced and that is done by pumping heat into the sample during the secondary drying phase. This phase is relatively short, lasting 1–2 h, but impor

tant for long-term viability. However, over drying of the bacteria can be detrimental as well. Once again, based on the idiosyncrasies of your lyophilizer and samples, the ideal time for secondary drying needs to be determined experimentally. Generally, raise the shelf temperature to 20 °C and dry for 2 h.

10. With the vacuum in place, stopper the vials using the stoppering plate/mechanism. Release the vacuum, remove the vials, and further secure the rubber bungs/stoppers with foil crimp seals. It is best to store the vials at 4 °C in the dark.
11. Test the freeze dried bacteria for viability as compared to the original culture. Additionally, monitor the stability/viability of the freeze dried cultures by testing at periodical intervals. A good protocol will yield nearly 100 % viable cells.

11.5.2.5.6. Freeze Drying Using a Manifold

1. Once bacteria have been dispensed into vials/tubes, freeze in a -80 °C freezer or equivalent. Flash freezing can be done in a dry ice/ethanol bath, but such samples tend to dry slower. Keep the samples frozen using dry ice, until they are connected to the manifold.
2. Turn on the lyophilizer and condenser/cold trap. The manifold valves should be turned off and the vacuum turned on. Allow the vacuum to pull down to 200 mtorr or less.
3. Expediently connect a vial/tube to the manifold and open the valve. The vacuum will immediately start the sublimation process and pull heat from the sample. In turn, hook up the remaining samples. The vacuum will increase each time a valve is open, but it should begin to lower immediately. If the vacuum does not drop after a tube is attached, there might be a leak in that connection thus shut that valve and proceed to the others.
4. Freeze drying with a manifold relies on ambient heat to drive the sublimation of the water. As the available water decreases, the temperature will gradually climb to ambient. This may take 2–3 h. Often frost that forms on the outside of the tubes will dissipate once the sample is done.
5. Using an acetylene torch (propane will work but it takes longer), seal each vial or tube. Wear safety glasses to protect your eyes from shattering glass and gloves (such as cotton gardening gloves) to protect your hands from the hot glass.
6. Sealed vials should be stored at 4 °C in the dark.
7. Test the freeze dried bacteria for viability as compared to the original culture. Additionally, monitor the stability/viability of the freeze dried cultures by testing at periodical days intervals up to 1 year. If viability starts to decline rapidly, modification of the protocol is probably necessary.

11.6 Cryopreservation

Cryopreservation refers to the storage of a living organism at ultralow temperature (-196°C) such that it can be revived and restored to the same living state as before it was stored. It is a very reliable method and is generally considered superior to other preservation methods. Bacteria preserved in liquid nitrogen normally show high survival rates and good strain stability during long-term storage. In liquid nitrogen storage of microorganisms, polypropylene cryotubes, glass vials, glass capillaries, and polypropylene straws are generally used.

Several factors can affect cell viability and stability during cryopreservation. During cryopreservation, dehydration of cells results and osmotic imbalance is created due to the changes in the concentration of salts and other metabolites. During the cooling process, rupture of the cellular membranes can also occur by the formation of large ice crystals. Successful preservation can be achieved by the use of cryoprotective agents (e.g., dimethylsulfoxide and glycerol), maintaining a controlled rate of cooling (about $1^{\circ}\text{C}/\text{min}$ to about -30°C), and an appropriate rewarming protocol (rapid thawing in a 37°C water bath which takes about 1 min for a glass ampoule and somewhat longer for a plastic vial). In practice, a relatively slow cooling rate can be easily obtained by keeping ampoules/vials in mechanical deep freezers for 1–2 h or in the neck of the liquid nitrogen storage unit for some minutes and then lowering containers into it. It is, however, not good practice to plunge cultures directly into liquid nitrogen, as the liquid nitrogen may seep into any imperfectly closed or sealed capillaries, ampoules, or vials containing the bacterial suspensions. On removal from storage, nitrogen (inside an ampoule) will virtually instantly change to the gaseous phase causing an explosion. For safety reasons it is thus recommended that cultures should be stored in the gas phase of liquid nitrogen.

While preparing cells for cryopreservation, several factors such as optimal growth conditions, physiological state of the cells (preferably from the late logarithmic to early stationary phase of growth), high cells density should be considered as these can affect cell viability after cryopreservation. After mixing, cell suspensions should be kept for equilibration with the cryoprotective agent. For harvesting, liquid cultures are centrifuged. However, vigorous pipetting and high-speed centrifugation should be avoided and cells should be handled gently. Viability assays should be performed on all cultures before and after cryopreservation to assure long-term viability. To assure purity, identity of the preserved cultures should be verified and after freezing cultures should be recharacterized to assure their stability. Safety

precautions should be observed when removing an ampoule from liquid nitrogen. The face shield, laboratory coat, and insulated gloves should be worn as protection against liquid nitrogen splash and exploding ampoules. The level of the liquid nitrogen in the containers should be checked preferably on a daily basis and maintained to a constant level, as any drop in liquid nitrogen level below a critical volume can result in damage due to the warming of the samples.

11.6.1. Materials

- Screw-capped plastic cryovials (2 ml)
- Screw-cap glass ampoules (10 × 30 mm) of 2 ml capacity (see *Note 6*)
- Liquid nitrogen storage tanks with canisters, racks, and canes
- Hungate tubes with septa
- Butyl rubber overflows tubes (5 mm) with Luer Lock adapters at both ends and long syringe needles (10–15 cm in length)
- Sterile gas-tight hypodermic Luer Lock syringes
- Cryoprotective agent (glycerol and dimethylsulfoxide)

11.6.2. Methods

1. Harvest cells from late log or early stationary growth and preferably in active phase of growth. Scrape cells from the growth surface if they are anchorage to media. Centrifuge broth or anchorage-independent cultures to obtain a cell pellet, if desired.
2. Prepare presterilized glycerol (10 %, w/v) or DMSO (5 %, v/v) in the concentration desired in fresh growth medium. When mixing with a suspension of cells, prepare the cryoprotective agents in twice the desired final concentration.
3. Centrifuge the culture for 30 min at $4,000\times g$ in the screw-cap bottles in which cultures are grown and remove the supernatant anaerobically under a stream of nitrogen gas using an overflow butyl rubber tube (5 mm) with Luer Lock adapters at both ends and fitted with long syringe needles. To obtain sterile nitrogen gas, sterile cotton filled syringe is attached to a conduit connected to the N₂ gas (99.99 %) cylinder.
4. Resuspend the pellet carefully in ice cold sterile DMSO solution (5 %, v/v). In case of halophilic strains or cells which do not form a pellet, a thick bacterial suspension (in growth medium) is mixed in the ratio 3:1 with ice cold sterile DMSO (20 %, v/v). For extreme halophilic strains, optimum salt concentration should be maintained after mixing cell suspension with the DMSO. The cells are allowed to equilibrate with the cryoprotectant (15 min for DMSO, 30 min for glycerol) in an ice bath.

5. Filling of ampoules and freezing: While equilibrating, an aliquot of 1.0–1.5 ml of cell suspension is dispensed in to each plastic cryovial or glass ampoule. For anaerobes using a sterile gas-tight syringe (5–10 ml), the ampoules are evacuated for anaerobiosis to facilitate filling. About 1 ml of thick cell suspension (equilibrated with the DMSO) is withdrawn with a 1 ml sterile oxygen-free syringe (already flushed with nitrogen gas) and injected into each ampoule. Immediately after the glass ampoules or cryotubes are clamped onto labeled aluminum canes, they are placed at -30°C for about 1 h or for few minutes in the gas phase of liquid nitrogen. The canes are then placed in canisters, racks, or drawers and frozen by direct immersion in liquid nitrogen or preferably in the gas phase of liquid nitrogen.
6. Revival of cultures: The frozen ampoule is removed from liquid nitrogen. For partial thawing, these are immediately immersed in the mini water bath at 37°C for a few seconds. After thawing, the outer surface of the ampoules is dried by wiping and plastic vials are wiped with alcohol-soaked gauze prior to opening. For aerobic bacteria, the screwcap glass vials can be opened and flame sterilized at the neck. The thawed contents of the ampoule/vial are immediately transferred to fresh growth medium to dilute the cryoprotectant, which otherwise is lethal at higher temperatures. For anaerobes, the septum of the glass vial is flame sterilized after putting a drop of alcohol and with a 1-ml oxygen-free syringe, a small volume (~ 0.05 ml) of inoculum is withdrawn and injected into 5–10 ml liquid growth medium. The rest of the cell suspension is immediately frozen again (a self-made wax block rack, chilled to -30°C is used for transportation to the liquid nitrogen container) in liquid nitrogen for later use. In this way, one vial can be used for several repeated retrievals or inoculations. The DMSO which is often toxic during growth is diluted 100–200 times in the culture medium to a noninhibitory concentration. The inoculated growth medium is incubated under appropriate growth conditions.
7. Estimation of viability counts: For aerobic bacteria, 0.5 ml of inocula is transferred to 4.5 ml of liquid growth medium and serial decimal dilutions are prepared. Plating and counting are done using standard methods. For the estimation of viable cell counts in anaerobic bacteria, 0.5 ml of inocula is transferred from the unfrozen (for cell counts before freezing) and from the thawed cell suspension (for cell counts after freezing) into prerduced 4.5 ml medium in screw-cap tubes and 6–8 serial decimal dilutions are prepared using oxygen-free syringes and incubation is done under appropriate conditions. Colony counts on agar plates can be performed in an anaerobic

glove box or anaerobic jars. In the case of viable colony counts in agar roll tubes or on plates, the number of colonies are counted from each dilution and average colony forming cells per sample are calculated.

11.7 Safety

Safety precautions must be considered when preserving microorganisms by freeze-drying, freezing, and storing at cryogenic temperatures.

1. Culture handling: When opening any microbial culture, frozen or freeze-dried cultures, take care to prevent dispersion of the ampoule contents. Open these preparations in a biological safety cabinet if possible and perform all work with hazardous cultures in a biological safety cabinet. There are varying degrees of pathogenicity among microorganisms. All laboratory personnel should be aware of the hazards posed by the cultures they are handling.
2. Cryogenic storage: Because of its extremely cold temperature, liquid nitrogen can be hazardous if improperly used. When handling liquid nitrogen, take precautions to protect your eyes, face, and skin from exposure to the liquid. Wear protective clothing, including a laboratory coat, gloves designed for handling material at cryogenic temperatures, and a face shield. To reduce your exposure to cryogenic temperatures, design inventory systems for storing frozen specimens that allow for easy retrieval and that minimize the time required to look for specimens. Prolonged exposure to cryogenic temperatures can lead to a loss of sensation in the hands that can only be recovered after warming. This loss of sensation can lead to a false sense of security regarding damage to tissues by the low temperatures. When the temperature in a liquid nitrogen unit becomes tolerable and working in the unit is no longer uncomfortable, the operator has reached a point where damage from the cryogenic temperatures is likely. When liquid nitrogen is used in confined and inadequately ventilated areas, the nitrogen can quickly displace the room air. Liquid nitrogen freezers should be located in well-ventilated areas, and special precautions should be taken during fill operations. In facilities with several liquid nitrogen freezers, an oxygen monitor should be installed to warn occupants of any deterioration in the air quality due to the nitrogen gas. Plastic screw-capped vials can present a hazard if stored directly in liquid nitrogen. Vials with an inadequate seal between the cap and the vial can fill with liquid nitrogen. Upon retrieval to warmer

temperatures the vials may explode violently or may spray liquid, potentially disseminating the contents of the vial. Likewise when opening plastic vials after thawing some dissemination of the contents may occur. Therefore, material in plastic ampoules should be stored in the vapor above the liquid nitrogen.

3. Freeze-drying: When freeze-drying microorganisms in vials or ampoules without cotton plugs or other bacteriological filters, the microorganisms can be carried from the container and contaminate the freeze-drying system. Microbial contamination can be found on the outside of the vial or ampoule and on parts of the freeze-drying system such as the condenser. A system should be designed to monitor the contamination level, and decontamination procedures should be implemented if necessary. Take care to properly treat freeze-dried cultures prior to disposal. To autoclave freeze-dried cultures, open the vial or ampoule to allow penetration of the steam. An alternative to autoclaving is to heat the preparations in a hot air oven at 180 °C for 4 h.

11.8 Notes

1. Usually, it is preferred to use tubes that are screw cap and have caps with O-rings. One option is to fill the tube 1/3 full with glass beads (4 mm) before sterilizing. When retrieving bacteria, remove or chip out one bead onto an agar plate and roll around to disperse the bacteria. This avoids thawing the entire stock.
2. It is vital to keep accurate records of lyophilized bacteria as well as using dependable techniques for labeling individual samples. It is recommended to use hard copy, freeze-drying log which can be used to record each batch of samples processed. The labeling of tubes/vials is also critical and should be done using a durable label. A clever labeling technique can be done by placing small, sterile paper labels inside the sample tube, which is subsequently sealed in the tube along with the sample.
3. Bacteria need a lyoprotectant which helps them survive the freeze-drying process. This medium can be very simple, such as 10 % skim milk, or complicated such as those that use animal sera. Good media have two main components: the lyoprotectant that stabilizes the cells when water is removed, and matrix agent that allows the entire sample to retain its shape during and after processing. Disaccharides such as sucrose and trehalose are excellent lyoprotectants. Matrix

forming additives, often referred to as excipients, include mannitol, BSA, serum, and skim milk.

4. Sucrose is a lyoprotectant and provides good viability, especially when compared to skim milk. However, bacteria preserved in sucrose must be kept cold during lyophilization to prevent melting and collapse of the sample.
5. The choice of vial or tube is very important for long-term survival of freeze-dried bacteria. Generally, vials/tubes used for lyophilizing bacteria are made of glass. Avoid usage of microfuge tubes for long-term storage as water can pass through plastic. Tubes, including ampoules, are the best containers for long-term storage of freeze-dried bacteria. These are most commonly attached to a manifold that holds multiple tubes. Tubes and ampoules are frozen using a freezer or dry ice bath and then quickly connected to the manifold before they melt. After the samples are dried, the neck of the tube or ampoule is sealed off using a propane or acetylene torch. Tubes and ampoules sealed under vacuum are impervious to moisture. The downside is that they require much more labor in comparison to vials. For freeze-drying in a shelf lyophilizer, vials are equipped with a stoppering plate. Vials are filled and then fitted with a stopper which has a notch that allows gas flow while sitting loosely in the vial opening. After freeze-drying, the stoppering plate is lowered and pushes the stoppers into the vial under vacuum. When the vacuum is released, atmospheric pressure secures the stopper and the vacuum within the vial. After removing from the lyophilizer, stoppers are further secured with an aluminum band which is crimped in place. Vials are very convenient and easy to use, but they can leak during long-term storage. For the short-term, they are very good. Although, any borosilicate glass test tube or tubing can be used for freeze-drying. Borosilicate glass is more difficult to seal than soda-lime tubes, but it is more durable. In using tubes, the culture medium is added to a sterile tube which is then loosely plugged with sterile glass wool. The sample is frozen, hooked to the vacuum, and processed. Once dry, a torch is used to seal the tube between the sample and vacuum manifold.
6. Ampoules are the easiest container to seal with a flame due to their design. As above, the cell suspension is added to a sterile ampoule, usually with a Pasteur pipette or very narrow micro-pipette tip. The ampoule is then loosely plugged with sterile glass wool. Following processing, the ampoule is sealed using a flame. As the ampoule has a very thin neck, it is much easier to seal. Ampoules can also be purchased prescored which makes cracking them open very easy as compared to tubes.

References

1. Cockell CS, Jones HL (2009) Advancing the case for microbial conservation. *Oryx* 43(4): 520–526
2. Kashyap PL, Sanghera GS, Wani SH, Shafi W, Kumar S, Srivastava AK, Arora DK (2011) Genes of microorganisms: paving way to tailor next generation fungal disease resistant crop plants. *Notulae Scientia Biologicae* 3(4):147–157
3. Day JG, Stacey GN (2007) Cryopreservation and freeze-drying protocols, 2nd edn. Humana Press Inc., Totowa, New Jersey, p 368
4. Nakasone KK, Peterson SW, Jong SC (2004) Preservation and distribution of fungal cultures. In: Gregory MM, Bills GF, Foster MS (eds) Biodiversity of fungi: inventory and monitoring methods. Elsevier Academic Press, San Diego, CA, pp 37–47
5. Smith D, Onions AHS (1994) The preservation and maintenance of living fungi, 2nd edn. CAB International, Wallingford Oxon
6. Espinel-Ingroff A, Montero D, Martin-Mazuelos E (2004) Long-term preservation of fungal isolates in commercially prepared cryogenic microbank vials. *J Clin Microbiol* 42:1257–1259
7. Nagai T, Tomioka K, Takeuchi K, Iida M, Kawada M, Sato T (2005) Evaluation of preservation techniques of microorganism resources in the MAFF genebank. *Jpn Agric Res Quart* 39:19–27
8. Stalpers JA, De Hoog A, Vlug IJ (1987) Improvement of the straw technique for the preservation of fungi in liquid nitrogen. *Mycologia* 79:82–89
9. Schipper MAA, Bekker-Holtman J (1976) Viability of lyophilized fungal cultures. *Antonie Van Leeuwenhoek* 42:325–328
10. Kirsop BE, Doyle A (1991) Maintenance of microorganisms and culture cells: a manual of laboratory methods, 2nd edn. Academic, London, p 308

Chapter 12

Microbes from Extreme Environment: Molecular Identification Procedures

Surajit Das

Abstract

Extremophiles are the organisms that inhabit in the extreme habitats like very high temperature, low temperature, high pressure, or cardinal conditions of pH, salinity, or adverse nutrient conditions. To survive in adverse conditions, these microorganisms develop certain unique mechanisms like secretion of certain enzymes or proteins creating a unique interest for biotechnological applications. Utilizing these microorganisms for the production of these useful substances at industrial scale, proper identification and characterization is required. Though certain conventional techniques are available for this purpose but they are of limited scope and it requires proper validations and modification. Molecular approach has broadened the limited range and DNA-based methods are emerging as the more reliable, simple, and inexpensive ways to identify and classify microorganism. Hence, in this chapter some identification techniques of these extreme organisms at molecular level have been described.

12.1 Introduction

Extreme environments include high temperature, pH, pressure, salt concentration, and low temperature, pH, nutrient concentration, and water availability, and also conditions having high levels of radiation, harmful heavy metals, and toxic compounds (organic solvents). Extremophiles are those microorganisms which inhabit in these environments. The discovery of *Thermus aquaticus* from Yellowstone National Park in 1965 by Professor Thomas Brock paved the way for thermophilic microbiology and in 1968 Professor K. Horikoshi coined the term alkaliphiles to describe the microorganisms thrive at alkaline pH. Thermophiles and alkaliphiles are now just two examples of extremophiles, microorganisms capable of living under extreme conditions. Others groups include acidophilic, barophilic, psychrophilic, and halophilic microorganisms.

These extremophiles are adapted to living at 100 °C in volcanic springs, at low temperatures in the cold polar seas, at high pressure in the deep sea, at very low and high pH values

(pH 0–1 or pH 10–11), or at very high salt concentrations (~35 %). The majority of the organisms that grow in these extreme environments belong to a group with distinct characteristics [1]. Carl Woese named this group archaea and postulated the archaea as the third domain of life on earth, different from bacteria and eukarya. In many cases microbial biocatalysts, especially of extremophiles, are superior to traditional catalysts, because they allow the performance of industrial processes even under harsh condition, under which conventional proteins are completely denatured. By virtue of their positive properties, stability, specificity, selectivity, and efficiency, enzymes already occupy a prominent position in modern biotechnology.

For many processes in the chemical and pharmaceutical industries, suitable microbial enzymes can be found that have the potential to optimize or even replace chemical processes. By using robust enzymes in biotechnical processes one is often able to better utilize raw materials, minimize pollutant emissions, and reduce energy consumption while simultaneously improving quality and purity of products, e.g., optically pure compounds. The additional benefits in performing industrial processes at high temperature include reduced risk of contamination, improved transfer rates, lower viscosity, and higher solubility of substrates. Extremophiles and their cell components, therefore, are expected to play an important role in the chemical, food, pharmaceutical, paper, and textile industries as well as environmental biotechnology.

Therefore, proper identification of the extremophilic strains is very much important. Conventional identification methods are available for only limited range of archaeal species. However, molecular approach to identification has broadened the limited range. DNA-based methods are emerging as the more reliable, simple, and inexpensive ways to identify and classify microorganism [2]. Extremophilic bacteria can be identified following Fluorescence in situ hybridization (FISH) with rRNA-targeted probes, Polymerase chain reaction-Denaturing gradient gel electrophoresis (PCR-DGGE) [3], terminal restriction fragment length polymorphism (T-RFLP) [4], 16S rRNA library construction, and 16S rDNA sequencing [5]. Two methods can be used for fungal ecology study: terminal restriction fragment length polymorphism (T-RFLP) analysis and denaturing gradient gel electrophoresis (PCR-DGGE) and for identification of actinomycetes strains, follow 16S rRNA library construction and 16S rDNA sequencing protocol and/or culture the colonies on solid medium, extract the genomic DNA, amplify 16S rDNA in PCR, sequence the 16S rRNA gene, compare the sequence from public database (e.g., NCBI or RDP), calculate the percentage of similarity and construct the phylogenetic tree [6].

12.2 Materials

12.2.1. Fluorescence *In Situ* Hybridization

12.2.1.1. For Fixation and Preparation of Sediment Samples

1. Microcentrifuge for 2 ml tubes
2. Vacuum pump
3. Ultrasonication probe
4. 2-ml screw-top microfuge tubes
5. 2-ml microfuge tubes
6. Plastic petri dishes (diameter, 5 cm)
7. White polycarbonate membrane filters (diameter, 25 mm, pore size 0.2 μm)
8. Cellulose nitrate support filters (diameter, 25 mm, pore size $\geq 0.45 \mu\text{m}$)
9. Filter tower for 25 mm membrane filters
10. 1 \times PBS
11. Ethanol
12. 4 % (w/v) formaldehyde solution

12.2.1.2. Hybridization on Filter Sections and Counterstaining

1. 2-ml microfuge tubes
2. 0.5-ml microfuge tubes
3. 50 ml polyethylene tubes and rack
4. Blotting paper
5. Razor blades
6. Plastic petri dishes
7. Microscopic slides + cover slips
8. CITIFLUOR mountant
9. VECTA SHIELD mountant
10. 1 M Tris/HCl, pH 7.4
11. Formamide
12. 0.5 M EDTA, pH 8
13. 10 % (w/v) sodium dodecyl sulfate (SDS)
14. 5 M NaCl solution
15. 4',6-Diamidino-2-Phenylindole (DAPI) dissolved in distilled H_2O , final concentration, 1 $\mu\text{g}/\text{ml}$
16. 80 % (v/v) ethanol

*12.2.1.3. Hybridization
with Horseradish
Peroxidase (HRP)-Labeled
Probes and Tyramide
Signal Amplification (TSA)*

1. 2-ml microfuge tubes
2. 0.5-ml microfuge tubes
3. 50 ml polyethylene tubes and rack
4. Blotting paper
5. Razor blades
6. Plastic petri dishes
7. Microscopic slides + cover slips
8. Parafilm
9. DAPI (4',6-diamino-2-phenylindole dihydrochloride)
10. 1 M Tris HCl, pH 7.4
11. Formamide
12. 0.5 M EDTA, pH 8
13. 10 % (w/v) sodium dodecyl sulfate (SDS)
14. 5 M NaCl solution

**12.2.2. PCR-
Denaturing Gradient
Gel Electrophoresis**

1. Formamide (deionized): Add 10 g of mixed bed resin (Sigma) to 100 ml formamide in an erlenmeyer and stir for 30–60 min. Decant or filter the formamide to separate it from the resin beads. Store the deionized formamide in volumes of 35 ml at -20°C for the preparation of the gel solution.
2. Acrylamide/bis-acrylamide stock solution (37.5:1; 40 % w/v): It can be purchased or prepare the solution from acrylamide powder (filter the solution and store at 4°C in a dark bottle).
3. DGGE acrylamide/bis-acrylamide solutions (for a gradient of 30–65 % of denaturant)
 - (a) 10 % acrylamide/0 % denaturant
 - 2.5 ml $50\times$ TAE
 - 62.5 ml acrylamide/bis-acryl 40 %
 - 185 ml H_2O distilled
 - (b) 8 % acrylamide/0 % denaturant
 - 2.5 ml $50\times$ TAE
 - 50 ml acrylamide/bis-acryl 40 %
 - 197.5 ml H_2O distilled
 - (c) 10 % acrylamide/100 % denaturant
 - 2.5 ml $50\times$ TAE
 - 62.5 ml acrylamide/bis-acryl 40 %

- 100 ml formamide
 - 105 g Urea + H₂O for 250 ml
- (d) 8 % acrylamide/100 % denaturant
- 2.5 ml 50× TAE
 - 50 ml acrylamide/bis-acryl 40 %
 - 100 ml formamide
 - 105 g Urea + H₂O for 250 ml

Note: These solutions have to be filtered through a 0.45-μm filter and stored at 4 °C.

4. Mixed gel solutions for a DGGE gel

30 % denaturing solution: 7.2 ml of 8 % acrylamide/100 % denaturant solution + 16.8 ml of a 8 % acrylamide/0 % denaturant solution [then, add 32 μl TEMED (*N,N,N',N'*-Tetramethylethylenediamine) and 75 μl 10 % APS (Ammonium persulfate)].

65 % denaturing solution: 15.6 ml of 10 % acrylamide/100 % denaturant solution + 8.4 ml of 10 % acrylamide/0 % denaturant solution (then, add 32 μl TEMED and 75 μl 10 % APS).

Stacking gel solution: Solution of 8 % acrylamide/0 % denaturant (10 ml per gel); add 20 μl TEMED and 40 μl 10 % APS.

Note: These solutions have to be degassed for 15 min under vacuum and stored at 4 °C in the dark. TEMED and 10 % APS are added before loading the gel.

5. 10× Gel loading solution: Bromophenol blue (0.025 g; 0.25 % w/v); xylene cyanole (0.025 g; 0.25 % w/v); glycerol (5 ml; 100 % v/v); water 5 ml

**12.2.3. Terminal
Restriction Fragment
Length Polymorphism**

1. Extraction buffer (100 mM Tris, 100 mM EDTA, 100 mM sodium phosphate buffer, pH 8.0)
2. Proteinase K (10 mg/ml)
3. Lysozyme (100 mg/ml)
4. Sodium Dodecyl Sulfate (10 %)
5. 5 M NaCl
6. 5 % CTAB (hexadecylmethylammonium bromide)
7. Isopropanol
8. 10 M ammonium acetate (pH 7.5)
9. PCR reagents
10. Deionized formamide

11. ROX-labeled GS500 internal size standard (Applied Biosystems)
12. Loading buffer
13. 5 % polyacrylamide gel

12.2.4. 16S rRNA Library Construction and 16S rDNA Sequencing

1. 0.12 M sodium phosphate buffer (PB buffer pH 8.0)
2. Lysis solution I (0.15 M NaCl, 0.1 M EDTA, pH 8.0, 10 mg lysozyme ml/1)
3. Lysis solution II (0.1 M NaCl, 0.5 M Tris-HCl, pH 8.0, 12 % SDS)
4. 5 M NaCl
5. 10 % TRITON-X100 in 0.7 M NaCl
6. CHCl₃
7. Isoamyl alcohol
8. 13 % PEG (polyethylene glycol dissolved in 1.6 M NaCl)
9. 70 % ethanol
10. Deionized H₂O

12.3 Methods

12.3.1. Fluorescence In Situ Hybridization

12.3.1.1. Fixation and Preparation of Sediment Samples

1. Suspend 0.5 ml of freshly collected sediment in 1.5 ml of 4 % formaldehyde solution in a 2-ml screw-top microfuge tube.
2. Fix for 1–24 h.
3. Centrifuge at 10,000 rpm for 5 min, pour off supernatant.
4. Add 1.5 ml of PBS and resuspend sample.
5. Centrifuge at 13,000 × *g* for 5 min, pour off supernatant.
6. Add 1.5 ml of a 1:1 mix of PBS/ethanol and store sample at –20 °C until further processing.
7. Resuspend sample and transfer 20–100 µl of aliquot to 500 µl of a 1:1 mix of PBS/ethanol in a 2-ml microfuge tube.
8. Sonicate aliquot for 20–30 s at low intensity using 1-s sonication pulses.
9. Place cellulose nitrate support filters beneath the membrane filters to improve the distribution of cells.
10. Add 15–20 µl of aliquot from the sonicated sample to 2 ml of distilled water and filter this volume onto the membrane filters.
11. Air-dry filtered preparations and store in petri dishes at –20 °C until hybridization.

Table 12.1
Composition of hybridization buffer

Formamide (%)	Formamide (μl)	NaCl 5M + Tris HCl 1M + SDS 10 %	H ₂ O (μl)
0	0	360 μl + 40 μl + 2 μl	1,598
5	100	360 μl + 40 μl + 2 μl	1,498
10	200	360 μl + 40 μl + 2 μl	1,398
15	300	360 μl + 40 μl + 2 μl	1,298
20	400	360 μl + 40 μl + 2 μl	1,198
25	500	360 μl + 40 μl + 2 μl	1,098
30	600	360 μl + 40 μl + 2 μl	998
35	700	360 μl + 40 μl + 2 μl	898
40	800	360 μl + 40 μl + 2 μl	798
45	900	360 μl + 40 μl + 2 μl	698
50	1,000	360 μl + 40 μl + 2 μl	598

12.3.1.2. FISH
Probe-EUB388

1. Specificity: Most bacteria and archaeobacteria
2. Target molecule: 16S rRNA
3. Position: 338–355; T_m : 55 °C; Formamide: 0–50 %
4. Sequence: 5'-GCT GCC TCC CGT AGG AGT-3'
5. Labeling: 5' labeling with Fluorescein-Isothiocyanate (FITC)
6. Prepare the probe solution: 2 μl probe + 18 μl Hybridization buffer/section

12.3.1.3. Preparation
of Hybridization Buffer
and Washing Buffer

1. Prepare the “Hybridization Buffer” of 0–50 % formamide in 2 ml-eppendorf tubes following the instructions (Table 12.1) and heating at 48 °C for 10 min.
2. Prepare the washing buffer of 0–50 % formamide in 2 ml-eppendorf tubes following the instructions (Table 12.2) and heating at 48 °C for 10 min.

12.3.1.4. FISH onto
Polycarbonate Filters

1. Filter the sample and cut in sections with a razor blade and number the each section with a pencil outside the filtered zone.
2. Cover a microscope slide with Parafilm and then put 10 μl of the “Probe solution” onto the slide.

Table 12.2

Composition of washing buffer

Formamide (%)	NaCl 5M + Tris HCl 1M + SDS 10 %	H ₂ O (μl)
0	360 μl + 40 μl + 2 μl	1,598
5	252 μl + 40 μl + 2 μl	1,498
10	180 μl + 40 μl + 2 μl	1,398
15	127.2 μl + 40 μl + 2 μl	1,298
20	86 μl + 40 μl + 2 μl	1,198
25	59.6 μl + 40 μl + 2 μl	1,098
30	40.8 μl + 40 μl + 2 μl	998
35	28 μl + 40 μl + 2 μl	898
40	18.4 μl + 40 μl + 2 μl	798
45	12 μl + 40 μl + 2 μl	698
50	7.2 μl + 40 μl + 2 μl	598

3. Leave the filter section onto the drop and then add 10 μl of “Probe solution” onto the filter section.
4. Make a dark humid chamber with a 50 ml-falcon tube with a folded piece of tissue inside and pour the rest of the “Hybridization buffer” onto the tissue.
5. Transfer the hybridization chamber immediately to the hybridization oven at the hybridization temperature for 1.5–3 h.
6. After that, wash the filter in the “Washing buffer” for 10 min at 48 °C.
7. Remove the excess washing buffer with distilled water in a Petri dish for 1–2 min in the dark.
8. Finally dry the filter section with filter paper and mount with coverslip.

12.3.1.5. Counterstaining with DAPI

1. Counterstain the filter section with DAPI (4',6-diamino-2-phenylindole dihydrochloride).
2. Place one drop of 1.5 μg/ml DAPI onto the filter section and then place the cover slip again onto the filter section.
3. Observe the sample directly by an epifluorescence microscope.

12.3.2. PCR-Denaturing Gradient Gel Electrophoresis

12.3.2.1. Extraction of DNA from Environmental Samples and PCR

1. Extract the genomic DNA following the methods described elsewhere in this manual.
2. Perform PCR amplification of the 16S rRNA fragment prior to DGGE using the following DGGE specific primers: F 5'CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG GCC TAC GGG AGG CAG CAG 3' and R 5'ACC GCG GCT GCT GG 3' [7].
3. Prepare the PCR reaction mixture in a volume of 50 µl containing 1.0 U of *Taq* polymerase, each primer at a concentration of 0.5 µM and 50 ng DNA template.
4. Amplification condition and PCR program is as follows:
 - (a) 94 °C for 12 min
 - (b) 94 °C for 1 min
 - (c) 50 °C for 1 min
 - (d) 72 °C for 2 min
 - (e) Go to step 2, 24 times
 - (f) 72 °C for 7.5 min
 - (g) 15 °C for ever
 - (h) End
5. Check the PCR amplifications on 1 % (w/v) agarose gel before DGGE analysis.

12.3.2.2. DGGE Conditions

1. Perform DGGE by loading the purified PCR product onto an 8 % (w/v) polyacrylamide (ratio of acrylamide to bisacrylamide, 37.5:1) gel with a gradient of 30–65 % denaturant (100 % denaturant contained 7 M urea and 40 % (vv) formamide).
2. Electrophorese the gels in 1× TAE buffer at 30 V for 15 min and then 130 V for 4.5 h at 60 °C.
3. Stain the gels in purified water (MilliQ or Millipore) containing ethidium bromide (0.5 mg/L) and destain twice in 0.5 × TAE for 15 min each.
4. Capture the images using an image analyzing system.

12.3.2.3. Extraction of DNA from Acrylamide Gels

1. Excise the central 1 mm² portion of strong DGGE bands using a razor blade and soak in 50 µl of purified water (MilliQ; Millipore) overnight at 4 °C
2. Use 1 µl supernatant as template in a PCR-DGGE reaction preparing the following reaction mix

- (a) 1 U of Advantage 2 polymerase mix
 - (b) 2 μ l 10 \times buffer
 - (c) 2 μ l MgCl₂ (2 mM)
 - (d) 1.5 μ l dNTP (1.5 μ M each)
 - (e) 10 pmol of each primer (341F-5'-CGC CCG CCG CGC GCG GCG GGC GG GGC GGG GGC CGG GGG GC CTA CGG GAG GCA GCA G-3' and 534R-5'ATT ACC GCG GCT GCT GG-3')
 - (f) 0.5 μ l formamide
 - (g) 1 μ l template DNA
 - (h) Double distilled water 25 μ l
3. Set the program accordingly
 - (a) 34 cycles
 - (b) 20 s at 96 °C
 - (c) 45 s 60 °C
 - (d) 45 s 72 °C
 - (e) 1 cycle
 - (f) 5 min at 72 °C
 - (g) Check the products in a 1 % agarose/1 \times TBE gel and then purify using the PCR clean-up centrifugal device

12.3.2.4. Sequencing of Purified PCR Products and Cloning

1. Sequence the PCR products derived from strong DGGE bands, free of minor bands directly using the purified PCR product.
2. Perform the sequencing reaction.
3. Clone the amplification products into either pCR II or pCR4-TOPO plasmid vectors.
4. Transform the ligated products into competent *Escherichia coli*.
5. Extract the DNA from the recombinant (white) colonies and use the aqueous DNA solution as DNA template in the PCR reaction.
6. Analyze the products by gel electrophoresis and digest with restriction enzyme EcoRI to check the presence of different cloned bands.
7. Select the high frequency sequences in clone libraries for sequence analysis and phylogenetic tree construction.

12.3.2.5. Phylogenetic Analyses

1. Analyze the percent similarity of DNA sequences through BLAST.
2. Perform the multiple sequence alignment using Clustal X software.
3. Construct the phylogenetic and molecular evolutionary trees to get the species affiliation using maximum parsimony, neighbor-joining, or maximum-likelihood analyses with bootstrap values.

12.3.3. Terminal Restriction Fragment Length Polymorphism

12.3.3.1. Isolation of Community DNA

1. Resuspend the soil sample (5 g) in 12 ml of extraction buffer.
2. Add 100 μ l proteinase K (10 mg/ml) and 180 μ l of lysozyme (100 mg/ml).
3. Shake the sample at 37 °C for 30 min followed by the addition of 3 ml of 10 % SDS, 4.5 ml of 5 M NaCl, and 1.5 ml of 5 % CTAB (hexadecylmethylammonium bromide)/1.5 M NaCl and incubate further at 65 °C for 15 min.
4. Freeze the samples in liquid nitrogen and subsequently thaw at 65 °C for a further 15 min. Repeat this freeze-thaw cycle twice.
5. Centrifuge at 4,200 rpm for 10 min at 48 °C.
6. Transfer the resulting supernatant in a fresh sterile tube.
7. Resuspend the remaining pellet by vortexing for 10 s in 4 ml of extraction buffer, 1 ml of 10 % SDS, and then incubate at 65 °C for 10 min.
8. Centrifuge at 4,200 rpm for 10 min at 48 °C.
9. Remove the supernatant and combine with that from the first centrifugation.
10. Add an equal volume of phenol:chloroform:isoamylalcohol (25:24:1) to the combined supernatant and mix by inversion.
11. Centrifuge the sample at 4,200 rpm for 10 min at 48 °C and transfer the aqueous upper layer to a fresh sterile tube.
12. Add an equal volume of chloroform:isoamylalcohol (24:1), mix by inversion, centrifuge at 4,200 rpm for 10 min.
13. Transfer the aqueous phase and add isopropanol (0.7 vol.) and 0.3 vol. of 10 M ammonium acetate (pH 7.5) at room temperature to the supernatant.
14. Centrifuge at 12,000 rpm for 30 min.
15. Discard the supernatant and resuspend the pellet in 1.5 ml of 70 % ethanol in an Eppendorf tube.

16. Centrifuge at 14,000 rpm for a further 15 min.
17. Discard the supernatant and the air-dry the pellet before resuspension in 250 ml of sterile distilled water.

12.3.3.2. PCR Amplification

1. Perform PCR using 2.5 U of Taq DNA polymerase with the Archaeal primers: 25 F (5' CYG GTT GAT CCT GCC RG 3') and 1,492 R (5' GGT TAC CTT GTT ACG ACT T 3') [8].
2. Amplification condition and PCR program is as follows:
 - (a) 96 °C for 5 min
 - (b) 95 °C for 15 s
 - (c) 49 °C for 30 s
 - (d) 72 °C for 1 min
 - (e) Go to step 2, 24 times
 - (f) 72 °C for 5 min
 - (g) 15 °C for ever
 - (h) End
3. Visualize PCR product after electrophoresis on 0.8 % agarose TAE gel.
4. Purify the PCR product by the Montage™ PCR centrifugal filter device (Millipore Corp., USA) and the DNA was eluted in a final volume of 50 µl.

12.3.3.3. T-RFLP Analysis

1. Digest the purified PCR product (10 µl) was with 20 U of either *AluI* or *HhaI* in a total volume of 15 ml at 37 °C for 3 h.
2. Mix this restriction digestion (2 ml) with 2 ml of deionized formamide, 0.5 ml of ROX-labeled GS500 internal size standard (Applied Biosystems), and 0.5 ml of loading buffer.
3. Denature each sample by heating at 95 °C for 5 min and immediately transferring to ice.
4. Perform electrophoresis from aliquots (1.5 ml) of each digested product in a 36-cm 5 % polyacrylamide gel containing 7 M urea for 6 h at 3,000 V.
5. Visualize the RFLPs by staining with SYBER green I dye and then photograph.
6. Determine the lengths of fluorescently labeled T-RFs by comparison with internal standards by using GeneScan software (version 2.1) (Applied Biosystems).

12.3.3.4. Species Richness, Community Signature, and Community Similarity

1. Estimate the similarity of communities by visual comparison of the electropherograms or by numerically analyzing the pattern of T-RFLPs in gel images with GelCompare software (Applied Maths, Belgium).

2. Generate a stacked band pattern of T-RFLPs for each community from the same terminus following the normalization (no background subtraction) and image analysis steps of the software.
3. Produce the similarity matrix for the pattern fragments in the samples by calculating Jaccard coefficient. (The Jaccard coefficient considered the presence or absence of bands and the number of T-RFs in common in communities, as well as the total number of T-RFLPs observed).
4. Obtain a similarity dendrogram by the unweighted pair group method using average linkages. (A value of 0 means that the community fingerprints are completely different from one another and a value of 1 indicates that they are identical.)

12.3.4. 16S rRNA Library Construction and 16S rDNA Sequencing

12.3.4.1. DNA Extraction and Purification

1. Add 5 g soil sample to a 50-ml tube containing 5 ml 0.12 M sodium phosphate buffer (PB buffer pH 8.0).
2. Vortex for 1 min, incubate at room temperature for 10 min, and centrifuge at $7,700 \times g$ for 10 min.
3. Decant the supernatant and add 5 ml PB buffer to the soil pellet.
4. Repeat once the procedure of vortexing, incubation, and centrifugation.
5. Resuspend the soil pellet in 8 ml lysis solution I (0.15 M NaCl, 0.1 M EDTA, pH 8.0, 10 mg lysozyme), mix and incubate at 37 °C for 1 h with occasional gentle mixing.
6. Add 8 ml lysis solution II (0.1 M NaCl, 0.5 M Tris-HCl, pH 8.0, 12 % SDS) and pass the soil suspension through two cycles of freezing at -40 °C for 20 min and thawing at 65 °C for 20 min and then centrifuge at $7,700 \times g$ for 10 min.
7. Mix the supernatant with 2.7 ml 5 M NaCl and 2.1 ml 10 % TRITON-X100 in 0.7 M NaCl and incubate for 10 min at 65 °C.
8. Add an equal volume of Chloroform:isoamyl alcohol (24:1) and centrifuge for 5 min at $3,000 \times g$.
9. Transfer the supernatant to a clean tube.
10. Add an equal volume of 13 % PEG (polyethylene glycol dissolved in 1.6 M NaCl) to the supernatant, incubate on ice for 30 min, and centrifuge at $12,000 \times g$ for 10 min.

11. Decant the supernatant and wash the pellet with 5 ml 70 % cold ethanol and air-dry.
12. Resuspend the pellet of crude DNA extracts in 500 μ l deionized H₂O.
13. Quantify the DNA yield by DNA fluorometer after 0.8 % (w/v) agarose gel electrophoresis.
14. Store the DNA at -20°C until required for PCR amplification.

12.3.4.2. PCR Amplification

1. Use the crude DNA as template for the amplification of archaeal 16S rRNA genes via PCR.
2. The reaction mixture (25 μ l) contains the following components:
 - (a) 100 ng of genomic DNA
 - (b) Primers at 0.5 μ M each. Archaeal primers are 25 F (5' CYG GTT GAT CCT GCC RG 3') and 1,492 R (5' GGT TAC CTT GTT ACG ACT T 3')
 - (c) dATP, dTTP, dCTP, and dGTP each at a concentration of 200 μ M
 - (d) mM MgCl
 - (e) 1 U *Taq* DNA polymerase
 - (f) PCR buffer
3. Amplification condition and PCR program is as follows:
 - (a) 94°C for 12 min
 - (b) 94°C for 1 min
 - (c) 50°C for 1 min
 - (d) 72°C for 2 min
 - (e) Go to step 2, 24 times
 - (f) 72°C for 7.5 min
 - (g) 15°C for ever
 - (h) End

12.3.4.3. Construction of 16S rRNA Library and 16S rDNA Sequencing

1. PCR products of 16S rRNA genes were cloned directly into the pCR12.1-TOPO1, pCR II, or pCR4-TOPO plasmid vectors.
2. Transform the ligated products into competent *E. coli*.
3. Detect the positive clones by the appearance of white colonies in LB plates containing 40 mg/ml of XGal and 50 μ g/ml ampicillin.

4. Extract the DNA from the recombinant (white) colonies and use the aqueous DNA solution as DNA template in the PCR reaction.
5. Analyze the products by gel electrophoresis and digest with restriction enzyme EcoRI to check the presence of different cloned bands.
6. Select the high frequency sequences in clone libraries for sequence analysis and phylogenetic tree construction.
7. Isolate the recombinant plasmids from overnight cultures by alkaline lysis and a restriction analysis with EcoRI to detect the insertion.
8. Analyze the products by gel electrophoresis and digest with restriction enzyme EcoRI to check the presence of different cloned bands by digestion patterns.
9. Select the high frequency sequences in clone libraries for sequence analysis and phylogenetic tree construction.

12.3.4.4. Molecular Identification and Phylogenetic Analysis

1. Check all the sequences for chimeras using the CHIMERA-CHECK online analysis program of the RDP-II database.
2. Determine taxonomic hierarchy of the sequences by BLAST of NCBI and RDP Analysis Tools of Ribosomal Database Project-II Release 9 (<http://rdp.cme.msu.edu/index.jsp>).
3. Perform multiple sequence alignment with CLUSTAL X selecting related sequences from the NCBI Taxonomy Homepage (Tax-Browser) and Ribosomal Database Project-II databases.
4. Construct the phylogenetic trees using the Neighbor-joining method with 500/1,000 Bootstrap replications.
5. Sequences that differ by <3 % are considered to belong to the same phylotype.
6. Sequences with similitude percentages below 95 % are assigned to the closest family.

References

1. Webster G, John Parkes R, Cragg BA, Newberry CJ, Weightman AJ, Fry JC (2006) Prokaryotic community composition and biogeochemical processes in deep seafloor sediments from the Peru Margin. *FEMS Microbiol Ecol* 58:65–85
2. Stackebrandt E, Goebel BM (1994) Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* 44:846–849
3. Muyzer G, De Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 59:695–700
4. Osborn MA, Moore ERB, Timmis KN (2000) An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environ Microbiol* 2:39–50

5. Paul JD (ed) (2001) Marine microbiology. Methods in microbiology, vol 30. Academic, San Diego, CA, p 666
6. Akkermans ADL, van Elsas JD, de Bruijn EJ (eds) (1996) Molecular microbial ecology manual. Kluwer, Dordrecht
7. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, New York
8. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds) (1999) Short protocols in molecular biology. Science Press, Beijing

Chapter 13

ELISA-Based Identification and Detection of Microbes

Jyoti Verma, Sangeeta Saxena, and Sunil G. Babu

Abstract

Enzyme-linked immunosorbent assay (ELISA) is an analytical technique to detect the presence of an antigen or antibody in a given sample. It shows wider applications in clinical diagnosis, in pathological studies, and in quality control studies. Virtually all microbial species have unique antigen(s), and such type of antigen(s) can be exploited as specific molecules of detection by ELISA. The variations in ELISA allow us to detect either antigen or antibody, identifying the different strains of microbes at a time and also in characterization of the epitope distribution on the microbial surface. Five types of variants have been developed in this assay: (1) Direct ELISA-use in the detection of antigen; (2) Indirect ELISA-use in the detection of antibody; (3) Sandwich ELISA-use in identification of different epitopes at a time; (4) Competitive ELISA-use in quantifying the antigen/antibody, and (5) Multiplex ELISA-use in identification of multiple antigens/antibodies at a time. Here, we discuss about different variants of ELISA and add detailed steps in performing these ELISA methods with their advantages and limitations.

13.1 Introduction

Enzyme-Linked Immunosorbent Assay (ELISA) is a highly sensitive immunological technique with high specificity for the detection of particular antigen (Ag) or antibody (Ab) in a sample using enzyme-linked antibodies [1, 2]. ELISA was developed in 1970 and became a popular technique in laboratory diagnosis. It provides an ideal system for dealing with a wide range of studies in Biology attributing to its flexibility, whereby reactants can be used in different combinations, either attached passively to a solid phase support or in the liquid phase. An ELISA is a five-step procedure: (1) coat the microtiter plate wells with antigen; (2) block all unbound sites to prevent false-positive results; (3) add antibody to the wells; (4) add anti-antibody conjugated to an enzyme; and (5) reaction of a substrate with the enzyme to produce a colored product, thus indicating a positive reaction. ELISA has been used as a diagnostic tool in medicine and plant pathology, as well as quality control check in various industries. ELISA combines the

specificity of antibodies with the sensitivity of simple enzyme assays, as the enzyme-coupled antibodies are used in this method. The conjugated/linked enzyme reacts when a chromogenic or fluorogenic substrate is added and produces a color/signal that can be read through the ELISA reader. Generally, ELISAs are performed in 96/12 well strip formats which permits high-throughput results. The bottom of each well is coated with antigen or antibody to be tested, and then by adding enzyme-conjugated specific antibody followed by suitable substrate, color developed is measured and quantitated. For colorimetric assays the popular enzymes conjugated to the antibodies are alkaline phosphatase, horseradish peroxidase, urease, and β -galactosidase, etc.

Serological typing is an immuno-based technique and plays an important role in identifying the bacteria, based on the difference of antigenic determinants expressed on their cell surface [3, 4]. These surface antigens include lipopolysaccharides, capsular polysaccharides, membrane proteins, and extracellular organelles, such as flagella and fimbriae [5]. At present, ELISA is the most established immunological technique, from which the indirect/sandwich ELISA formats are the most commonly used ELISA formats for the detection of pathogens [6].

ELISA techniques have been developed for the detection of whole cell antigen targets or products for pathogens such as *Bacillus cereus* [7], *Campylobacter* spp. [8], *Escherichia coli* [9], and *Salmonella* spp. [10] from foods. Obst et al. [11] developed an ELISA using a monoclonal antibody against the enterobacterial common antigen (ECA), a lipopolysaccharide linked within the outer cell membrane of Enterobacteriaceae.

The use of immunological methods for the detection of specific microorganisms is a rapid and simple technique, the accuracy of which mainly depends on the specificity of the antibody [12]. One way to increase specificity is to select monoclonal antibodies which are highly specific in their action against a specific epitope. However, since an epitope can be present in more than one antigenic agent, a rigorous specificity testing of the monoclonal antibodies synthesized with closely and distantly related bacterial strains must precede the routine testing of environmental samples.

13.1.1. Principle

The basic principle of ELISA is to use an enzyme attached to antibody to detect either antigen or antibody in an immunoassay to allow quantification through the development of color after the addition of a suitable chromogenic substrate. ELISA is used to target the antigen with high specific antibodies which are conjugated to the enzymes. These enzyme-conjugated antibodies are allowed to bind to the respective antigens, and then after the addition of the suitable chromogenic substrate to this (Ag–Ab–E complex), color will be produced due to the enzymatic activity.

The enzymatic reaction stopped and the color developed will be detected spectrophotometrically by ELISA reader.

13.1.2. Types of ELISA

The ELISA method is a benchmark for quantitation of pathological antigens and there are indeed many variations to this method. ELISAs are adaptable to high-throughput screening because results are rapid, consistent, and relatively easy to analyze. ELISAs can be performed with a number of modifications to the basic procedure. Five types of ELISA formats have been developed, allowing us to estimate the qualitative and quantitative measurement of either antigen or antibody. They are

1. Direct ELISA
2. Indirect ELISA
3. Sandwich ELISA
4. Competitive ELISA
5. Multiplex ELISA

13.1.2.1. Direct ELISA

Direct ELISA is a method for detecting and measuring a particular antigen in a sample using a specific antibody conjugated to the enzyme. As the name indicates, the antigen is directly recognized and identified by antibody which itself is conjugated to the enzyme. Briefly, any microbial antigen(s) in purified form or complex form are coated on the surface of the microtiter well and these wells are allowed to incubate with the specific enzyme-conjugated antibodies, and after addition of the suitable substrate color will be developed. The development of the color shows the presence of antigen and the intensity of the color denotes the quantity of Antigen. The protocol of direct ELISA is illustrated in Fig. 13.1.

13.1.2.1.1. Advantages

It is relatively quick because only one antibody and fewer steps are used. It also eliminates the cross-reactivity of secondary antibody.

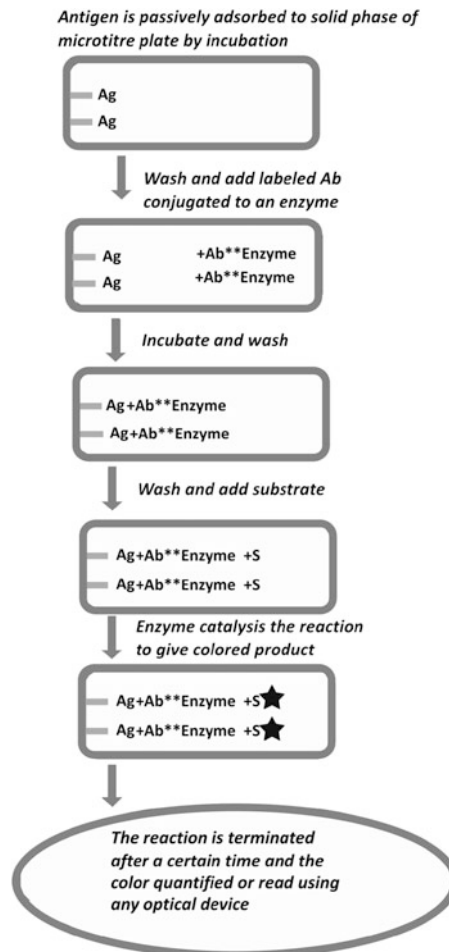
13.1.2.1.2. Limitations

The labeling of every primary antibody is a time-consuming and expensive proposition. Immunoreactivity of the primary antibody might be adversely affected by labeling with enzymes or tags. Further, certain antibodies are unsuitable for labeling. There is no flexibility in the choice of primary antibody label from one experiment to another and a minimal signal amplification in contrast to the methods that use secondary antibody labeling.

13.1.2.2. Indirect ELISA

The indirect ELISA is a two-step method using labeled secondary antibody for detection. First a primary antibody detects the antigen and then a secondary labeled antibody binds to the primary antibody that is used for color development. Briefly, sample containing primary antibody (Ab_1) is added to an antigen-coated

Fig. 13.1 Direct ELISA



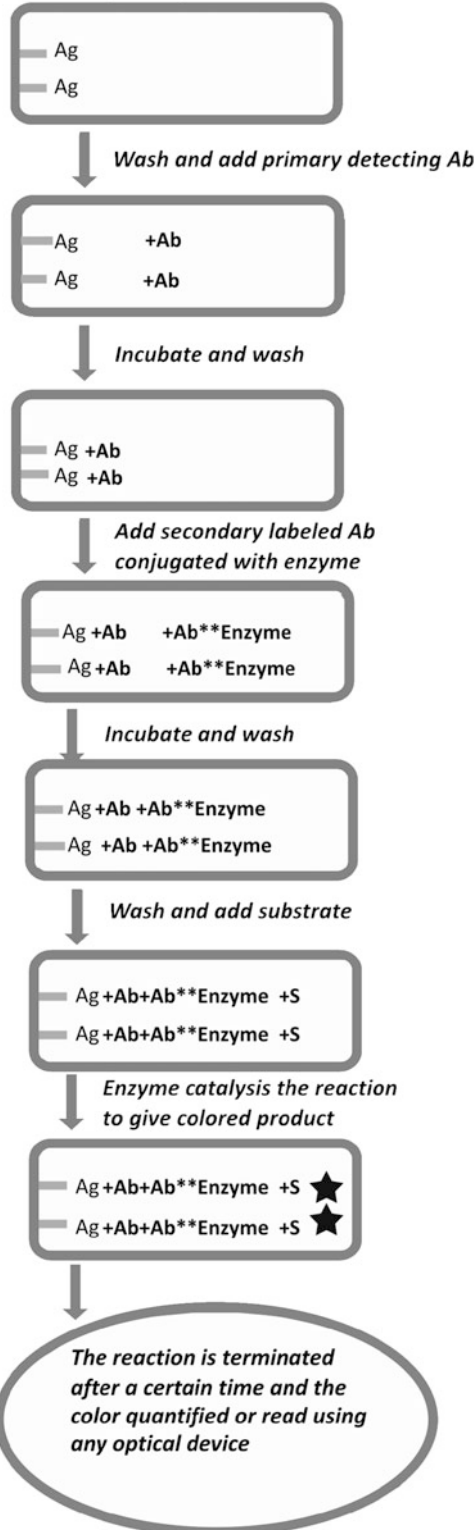
microtiter well and allowed to react with the antigen attached to the well. After washing, the enzyme-conjugated secondary antibody (Ab₂) which is specific to the isotype of primary antibody is added. After removing any excess antibody by washing, the addition of the substrate yields color, which can be measured through ELISA reader. The protocol of indirect ELISA is illustrated in Fig. 13.2. The indirect ELISA is a method of choice to detect the presence of serum antibodies against HIV.

13.1.2.2.3. Advantages

It offers the advantage that any number of antisera can be examined for binding to a given antigen using a single anti-species conjugate. This property has been heavily exploited in diagnostic applications, particularly when examining (screening) large numbers of samples. The signal amplification is also improved due to the use of secondary antibodies.

Fig. 13.2 Indirect ELISA

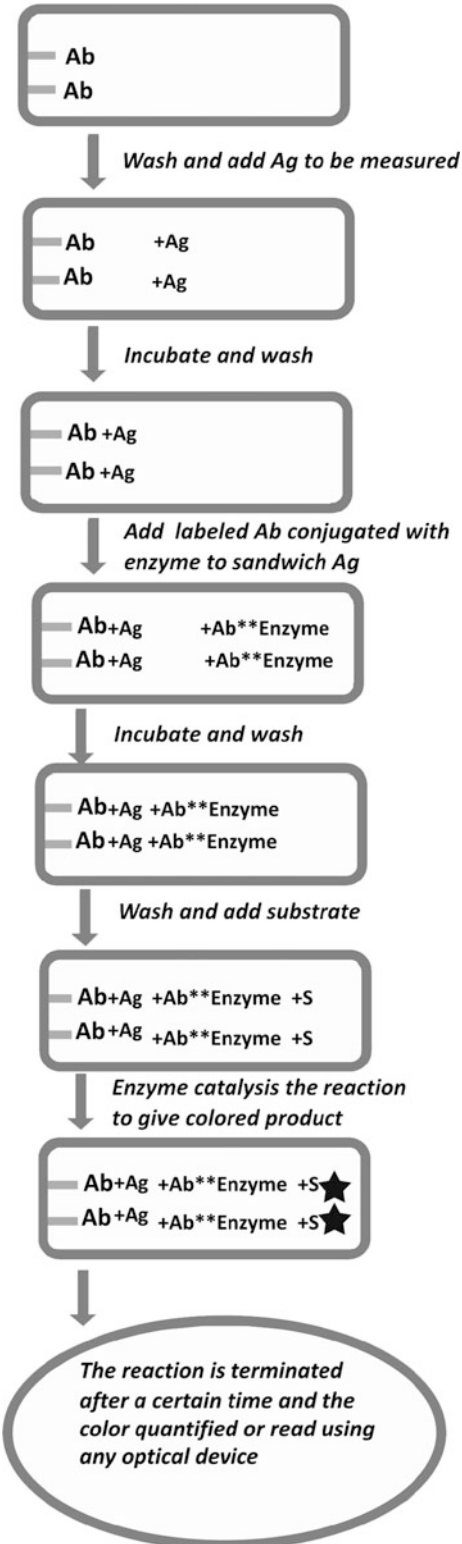
Antigen is passively adsorbed to solid phase of microtitre plate by incubation



- 13.1.2.2.4. Limitations One limitation with this type is that cross-reactivity may occur with secondary antibody resulting in nonspecific signal. An extra incubation step is required in the procedure.
- 13.1.2.3. Sandwich ELISA The sandwich ELISA measures the amount of antigen which is sandwiched between two layers of antibodies. Briefly, an antibody (the “capture” antibody) is allowed to bind to a solid phase, i.e., bottom of a microtiter well. Then a test sample containing an antigen is added and allowed to react with the bound antibody. After washing the well, a labeled second antibody (the “detection” antibody) which is specific to another epitope is added and is allowed to bind to the antigen, thus completing the “sandwich”. After the removal of any free antibody, substrate is added and the color can be measured. The basic protocol of sandwich ELISA is illustrated in Fig. 13.3.
- 13.1.2.3.5. Advantages It offers fast and accurate detection of antigen concentration in an unknown sample. Especially the antigen does not need to be purified prior to use. For example the strain identification of microorganisms or pathogens is generally done with this assay during epidemics. It is useful for quantization of antigens when antigen concentration is low and/or has higher concentrations of contaminating protein(s). Antigens need not be purified prior to use. This method gives reproducible results with high sensitivity.
- 13.1.2.3.6. Limitations One major disadvantage is that not all antibodies can be used. The antigens to be measured must contain at least two antigenic sites, capable of binding to the antibody as two antibodies act in the sandwich. For this reason, sandwich assays are restricted to the quantization of multivalent antigens such as proteins or polysaccharides.
- 13.1.2.4. Competitive ELISA This type of ELISA is used to quantify antigen using competitive method. Briefly, the antigen to be measured is incubated with controlled amount of antibody to form antigen–antibody complex. Then this antigen–antibody complex is added to an antigen-coated microtiter well. The more antigens present in the sample, the less free antibody will be available to bind to the antigen coated well. Now addition to the enzyme-conjugated secondary antibody specific to the primary antibody gives the amount of primary antibody bound to the well, which in turn gives the amount of antigen present in the sample. The protocol of competitive ELISA is illustrated in Fig. 13.4. In this type of ELISA, the higher the antigen concentration in the sample, the weaker the final signal, as less free primary Abs are available to bind to antigen on plate due to the competition for binding the antigen.

Fig. 13.3 Sandwich ELISA

Ab is passively adsorbed to solid phase of microtitre plate by incubation



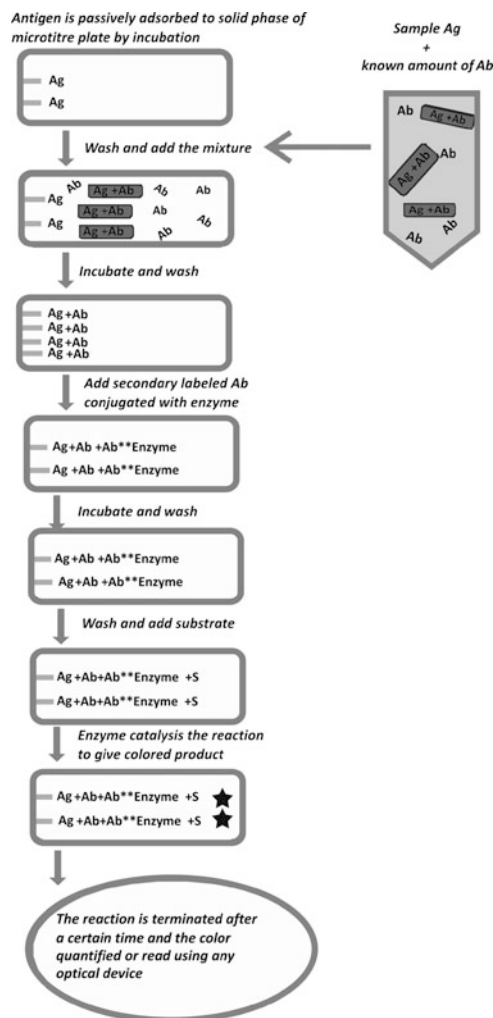


Fig. 13.4 Competitive ELISA

- 13.1.2.4.7. Advantages

The main advantage is the use of non-purified primary antibodies.
- 13.1.2.4.8. Limitations

Higher the original antigen concentration, the weaker the eventual signal. The more antigens in the sample, the less labeled antigen is retained in the well and the weaker the signal for some competitive ELISA.
- 13.1.2.5. Multiplex ELISA

The sandwich ELISA allows detecting the two different epitopes present on the antigen, suggesting the possibility of detection of multiple epitopes present on the antigen or sample(s) in the microtiter plate. The logical progression of the ELISA toward a protein array format which allows simultaneous detection of multiple antigens at multiple array addresses within a single microtiter well leads to the development of multiplex ELISA. Multiplexing in assays simply refers to the ability to output multiple readings

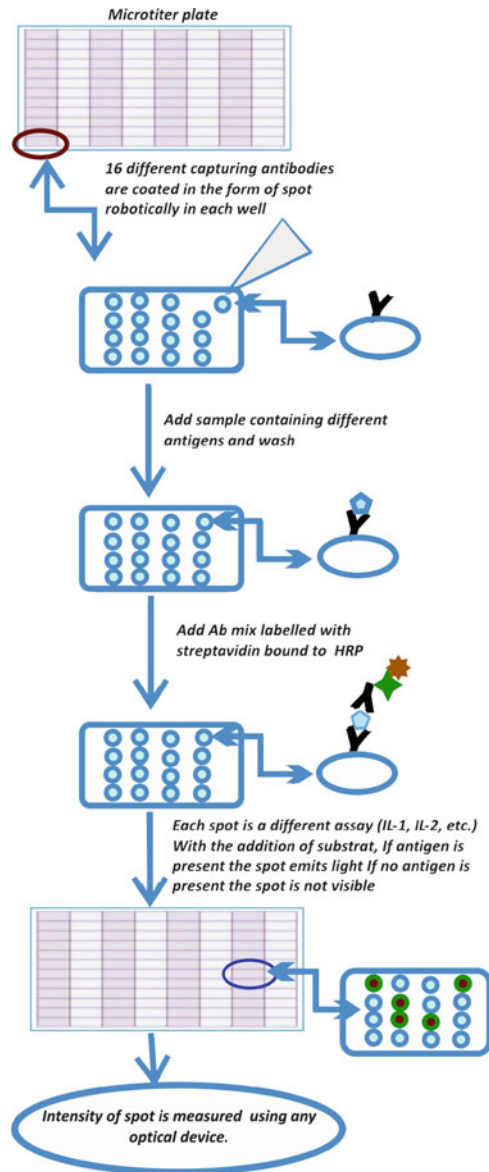


Fig. 13.5 Multiplex ELISA

from a single sample. Generally, multiplexing can be achieved through antibody array, where different primary antibodies are printed on glass plate to capture their respective antigens in given samples such as cell lysate, tissue extract, etc., allowing a single microtiter well to capture different antigens at a time. The protocol of multiplex ELISA is illustrated in Fig. 13.5. These different antigens captured are detected by direct, indirect, sandwich, or competitive ELISA depending upon antibody array technique used.

13.1.2.5.9. Advantages

Compared with the traditional ELISA, the multiplex arrays have a number of advantages including (a) high-throughput multiplex

analysis (Up to 25 Assays/well), (b) requirement of less sample volume, (c) efficiency in terms of time and cost, (d) ability to evaluate the levels of a given molecule in the context of multiple others, (e) ability to perform repeated measurements of the same formats in the same subjects under the same experimental assay conditions, and (f) the reliability of the detection of different proteins across a broad dynamic range of concentrations.

13.1.2.5.10. Limitations

In spite of the advantages, caution is necessary when considering the application of multiplex arrays. It requires experience/expertise to perform. Multiplex assays, by their very nature, involve potential interactions between multiple different antibodies and antigens in the sample/assay solution. One cannot assume that a reliable uniplex assay can just be simply added to a functioning multiplex assay. Non-reactivity to all other antibodies must first be established and the lowest amount possible must be used to minimize such cross-reactions.

The advent in the production of monoclonal/polyclonal antibodies, enzyme/fluorescent/chemiluminiscent conjugation to these antibodies, and development of solid phase immobilization of Ags made ELISA a versatile technique. Further, the variations developed in basic ELISA format made it most sensitive, handy, and robust, done within a shorter time, both easily and economically with reproducible results. The best results have been obtained with the sandwich format, utilizing highly purified, pre-matched capture and detector antibodies.

13.1.2.6. Applications

1. ELISA is a preferable technique than other immunological techniques because of its sensitivity (~ 0.0001 – $0.01 \mu\text{g ab/ml}$; can be increased to ~ 0.00001 – $0.01 \mu\text{g ab/ml}$) and is used in the detection and quantification of substances like peptides, proteins, antibodies, hormones, haptens, drugs, and their metabolites and of potential food allergens (milk, peanuts, walnuts, almonds, and eggs) which are more applied nowadays to check GM food.
2. Indirect ELISA is the most common and widely used technique in clinical diagnostics where it is used for detection of serum antibody concentrations, e.g., detection for the presence of antibodies in blood sample for past exposure to disease and outbreaks, such as Lyme disease, trichinosis, HIV, bird flu, etc.
3. Sandwich ELISA allows identifying the different strains of pathogen where the pathogen/antigen is limited.
4. Competitive assays are often used when the antigen to be measured is small and has only one epitope, or antibody binding site.

13.2 Materials

96-Well Microtiter Plates; ELISA Reader; Microplate shaker; Multichannel pipettor; Micropipettes and Eppendorf Tubes; Multiplex ELISA-based systems.

13.2.1. Buffers

13.2.1.1. Coating Buffer

1. (0.1 M Carbonate bicarbonate buffer, pH 9.6): 1 l
 - (a) 3.03 g Na_2CO_3
 - (b) 6.0 g NaHCO_3
 - (c) 1,000 ml distilled water
 - (d) Store at 4 °C
2. Phosphate buffer Saline (PBS), pH 7.4: 1 l
 - (a) 2.32 g Na_2HPO_4
 - (b) 0.2 g KCl
 - (c) 0.2 g K_3PO_4
 - (d) 8.0 g NaCl (500 ml distilled water)
 - (e) Store at 4 °C
3. Blocking buffer solution: 1 l
 - (a) 1,000 ml PBS buffer
 - (b) 0.5 ml Tween 20
 - (c) 10 g BSA
 - (d) Store at 4 °C
 - (e) (1 % BSA, serum, nonfat dry milk, casein, gelatin in PBS)
4. Washing solution: 1 l
 - (a) [PBS or Tris-buffered saline (pH 7.4) with detergent such as 0.05 % (v/v) Tween 20]
 - (b) 0.5 ml Tween 20
 - (c) 999.5 ml PBS buffer
 - (d) Store at 4 °C
5. Antibody dilution buffer
Primary and secondary antibody should be diluted in 1× blocking solution.

13.2.2. Reagents

Purified antigens/antibodies, Enzyme-conjugated (alkaline phosphatase, Horseradish peroxidase) antibodies (Primary/Secondary), PNPP (*p*-Nitrophenyl phosphate, Disodium salt), ABTS (2,2'-Azinobis [3-ethyl benzothiazoline-6-sulfonic acid]-diammonium salt), OPD (*O*-phenylenediamine dihydrochloride),

TMB (3,3',5,5'-tetramethyl benzidine), Hydrogen peroxide, Tween 20, and above-mentioned chemicals for the preparation of buffers.

13.3 Methods

13.3.1. Direct ELISA

13.3.1.1. Coating

1. Dilute the antigen with coating buffer and coat wells of ELISA plate 100 µl/well (see Notes 1).
2. Cover the plate with plastic and incubate 2 h at room temperature or overnight at 4 °C.
3. The antigen coating solution from the wells of plate by flicking the plate over sink. Remove and wash the plate thrice by filling the wells with 200 µl of washing buffer/PBS (see Notes 2).

13.3.1.2. Blocking

1. Block the nonspecific protein binding sites in the coated wells by adding 200 µl blocking buffer/5 % serum in PBS/well (see Notes 3).
2. Cover the plate with plastic and incubate 2 h at room temperature or overnight at 4 °C.
3. Remove the blocking buffer from the wells by flicking the plate over sink and wash the plate thrice with PBS.

13.3.1.3. Incubation

1. Add 100 µl of the antibody conjugated to enzyme in blocking buffer to each well (see Notes 4).
2. Cover the plate with plastic and incubate 2 h at room temperature.
3. Remove the enzyme-conjugated antibody solution from the wells by flicking over the sink and wash the plate four times with PBS/washing buffer.

13.3.1.4. Detection

1. Add 100 µl or 50 µl of substrate (for ALP or HRP accordingly) per well with multichannel pipette (see Notes 5).
2. After sufficient color development, add 50 µl of stopping buffer (8N H₂SO₄) (see Notes 6).
3. Read the absorbance of each well at 492 nm with a Microplate reader (see Notes 5).

13.3.2. Indirect ELISA

13.3.2.1. Coating

1. Dilute the antigen with coating buffer and coat wells of ELISA plate 100 µl/well (see Notes 1).

2. Cover the plate with plastic and incubate 2 h at room temperature or overnight at 4 °C.
3. Remove the antigen coating solution from the wells of plate by flicking the plate over sink and wash the plate thrice by filling the wells with 200 µl of washing buffer/PBS.

13.3.2.2. Blocking

1. Block the nonspecific protein binding sites in the coated wells by adding 200 µl blocking buffer/5 % serum in PBS/well.
2. Cover the plate with plastic and incubate 2 h at room temperature or overnight at 4 °C.
3. Remove the blocking buffer from the wells by flicking the plate over sink and wash the plate thrice with PBS.

13.3.2.3. Incubation

1. Add 100 µl of primary antibody or antiserum diluted in blocking buffer to each well (see Notes 4).
2. Cover the plate with plastic and incubate 2 h at room temperature or overnight at 4 °C.
3. Remove the primary antibody solution from the wells by flicking over the sink and wash the plate at least thrice with PBS/washing buffer.
4. Dilute the enzyme-conjugated secondary antibody with blocking buffer and add 100 µl to each well of the plate.
5. Cover the plate with plastic and incubate for 2 h at room temperature.
6. Remove the labeled secondary antibody solution from the wells by flicking over the sink and wash the plate at least five times with PBS/washing buffer.

13.3.2.4. Detection

1. Add 100 µl or 50 µl of substrate (for ALP or HRP accordingly) per well with multichannel pipette (see Notes 5).
2. After sufficient color development, add 50 µl of stopping buffer (8N H₂SO₄) (see Notes 6).
3. Read the absorbance of each well at 492 nm with a Microplate reader (see Notes 5).

13.3.3. Sandwich ELISA

13.3.3.1. Coating

1. Dilute the antibody with coating buffer and coat wells of ELISA plate 100 µl/well (see Notes 1a).
2. Cover the plate with plastic and incubate 2 h at room temperature or overnight at 4 °C.
3. Remove the antibody solution from the wells of plate by flicking the plate over sink and wash the plate thrice by filling the wells with 200 µl of washing buffer/PBS.

13.3.3.2. Blocking

1. Block the nonspecific protein binding sites in the coated wells by adding 200 μ l blocking buffer/5 % serum in PBS/well.
2. Cover the plate with plastic and incubate at 4 °C overnight or 2 h at room temperature.
3. Remove the blocking buffer from the wells by flicking the plate over sink and wash the plate thrice with PBS.

13.3.3.3. Standard and Samples Incubation

1. Dilute the standards/samples with blocking buffer and transfer 100 μ l to each well.
2. Cover the plate with plastic and incubate at 4 °C overnight or 2 h at room temperature.
3. Remove the standard/samples solution from the wells by flicking the plate over sink and wash the plate thrice with PBS.

13.3.3.4. Incubation with HRP-Conjugated Antibody

1. Dilute the labeled antibody (0.25–2 μ g/ml) in blocking buffer and add 100 μ l in each well.
2. Cover the plate with plastic and incubate for 1 h at room temperature.
3. Remove the blocking buffer from the wells by flicking the plate over sink and wash the plate thrice with PBS.

13.3.3.5. Detection

1. Add 100 μ l or 50 μ l of substrate (for ALP or HRP accordingly) solution per well.
2. After sufficient color development, add 50 μ l of stopping buffer (8N H₂SO₄).
3. Read the absorbance of each well at 492 nm with a Microplate reader.

13.3.4. Competitive ELISA**13.3.4.1. Coating**

1. Dilute the antibody with coating buffer and coat wells of ELISA plate 100 μ l/well (see Notes 1a).
2. Cover the plate with plastic and incubate 2 h at room temperature or overnight at 4 °C.
3. Remove the antibody solution from the wells of plate by flicking the plate over sink and wash the plate thrice by filling the wells with 200 μ l of washing buffer/PBS.

13.3.4.2. Blocking

1. Block the nonspecific protein binding sites in the coated wells by adding 200 μ l blocking buffer/5 % serum in PBS/well.
2. Cover the plate with plastic and incubate at 4 °C overnight or 2 h at room temperature.
3. Remove the blocking buffer from the wells by flicking the plate over sink and wash the plate thrice with PBS.

13.3.4.3. *Competitive Incubation*

1. Dilute the standards/samples with blocking buffer and dilute the enzyme-conjugated antigen in the blocking buffer.
2. Mix the standard/sample and HRP-conjugated antigen together and add 100 μ l of the diluted mixture to the wells.
3. Cover the plate with plastic and incubate at 4 °C overnight or 2 h at room temperature.
4. Remove the mixture solution from the wells by flicking the plate over sink and wash the plate thrice with PBS.

13.3.4.4. *Detection*

1. Add 100 μ l or 50 μ l of substrate (for ALP or HRP accordingly) solution per well.
2. After sufficient color development, add 50 μ l of stopping buffer (8N H₂SO₄).
3. Read the absorbance of each well at 492 nm with a Microplate reader.

13.3.5. **Multiplex ELISA**

13.3.5.1. *Coating*

1. Coat the wells of a microtiter plate with 100 μ l of antigen of concentration 0.5–10 μ g/ml in PBS or coating buffer and incubate for 2 h at room temperature.
2. Remove the coating solution and wash the plate (see Notes 1 and 1a).

13.3.5.2. *Blocking*

1. Block the wells by adding blocking solution (BSA/PBS) per each well.
2. Cover the plate with plastic and incubate for 2 h at room temperature or overnight at 4 °C and wash the plate (see Notes 3).

13.3.5.3. *Incubation with the Antibody*

1. Add the antibody conjugated to enzyme and cover the plate with plastic and incubate for 2 h at room temperature or overnight at 4 °C.
2. Wash the plate (see Notes 2).

13.3.5.4. *Detection or Color Development*

1. Fill the substrate solution per well with a multichannel pipette or a multichannel pipette.
2. Read the absorbance of each well with a plate reader after color development (see Notes 5 and 6).

13.3.5.5. *Analysis of Data*

1. Prepare a standard curve from the data produced from the serial dilutions with concentration on the X-axis vs. absorbance on the Y-axis.
2. Interpolate the concentration of the sample from this standard curve (see Notes 7).

13.4 Notes

13.4.1. Coating

1. Antigen or antibody should be diluted in coating buffer to immobilize them to the wells. Dilution is made according to the requirement; the concentration of coated antigen ranges from 1 to 10 $\mu\text{g}/\text{ml}$. Test samples containing pure antigen are usually pipetted onto the plate at $<2 \mu\text{g}/\text{ml}$. Antigen protein concentration should not be over 20 $\mu\text{g}/\text{ml}$, as this will saturate most of the available sites on the microtitre plate.
 - (a) The concentration of coated antibody ranges from 0.5 to 10 $\mu\text{g}/\text{ml}$
 - (b) Avoid contamination so be very careful and avoid spill-over from adjacent wells

13.4.2. Washing

1. Remove the coating solution and wash the plate thrice by filling the wells with 200 μl PBS. The solutions or washes are removed by flicking the plate over a sink. The remaining drops are removed by patting the plate on a blotting paper.
 - (a) Make sure that all incubations are carried out in 100 % humid conditions. Do not allow the plate to dry out between intermediate washing and incubation period. Always refrigerate plates in sealed bags with a desiccant to maintain stability and moisture content.

13.4.3. Blocking

1. Block the remaining protein binding sites in the coated wells by adding 200 μl blocking buffer and 5 % nonfat dry milk/PBS per well. Alternative blocking reagents include BSA.
 - (a) Although 2 h is usually enough to obtain a strong signal, if a weak signal is obtained, stronger staining will often observed when incubated overnight at 4 $^{\circ}\text{C}$.

13.4.4. Incubation

1. The concentration of incubated antibody is based on the manufacturer's instructions and is prepared immediately before use.

13.4.5. Detection

1. Four most popular enzymes used in ELISA are Alkaline phosphatase (ALP), Horseradish peroxidase (HRP), Urease, and β -galactosidase with respect to the substrate. For example, in case of alkaline phosphatase the substrate is *p*-nitrophenyl-phosphate (*p*NPP), and for peroxidases, 2,2'-azino-bis (3-ethylbenzothiazoline-6-sulfonic acid) (ABTS), *o*-phenylenediamine (OPD), and 3,3',5,5'-tetramethylbenzidine base (TMB) are used as substrates. For most applications *p*NPP (*p*-Nitrophenyl-phosphate) is the most widely used substrate for alkaline phosphatase and the yellow color of nitrophenol can be measured at 405 nm after 15–30 min incubation at

room temperature. This reaction can be stopped by adding equal volume of 0.75 M NaOH. The most commonly used substrate for HRP is hydrogen peroxide. Cleavage of hydrogen peroxide is coupled to oxidation of a hydrogen donor which changes color during reaction. Others are OPD, a light sensitive substrate kept and stored in the dark and measured at 492 nm, and ABTS (2,2'-azino-di-[3-ethyl-benzothiazoline-6 sulfonic acid] diammonium salt). The end product is green and the optical density can be measured at 416 nm.

- (a) ALP and HRP enzyme substrates are toxic if inhaled or swallowed. Avoid contact with skin.
 - (b) Sodium azide is an inhibitor of horseradish peroxidase. Do not include sodium azide in buffers or wash solutions, if an HRP-labeled conjugate will be used for detection.
 - (c) Avoid contact with the acidic Stop Solution and Substrate Solution, which contains hydrogen peroxide.
2. Generally 10 min is enough for color development for HRP conjugates OPD-H₂O₂, when the reaction is stopped by 8N H₂SO₄
 - (a) The antibody-enzyme conjugate cleaves the reagent and a color develops, and even a small amount of bound enzyme, if given enough time, produces more color; hence the reaction needs to be terminated by using weak acid. Otherwise all samples would yield the same absorbance and would be rendered indistinguishable. After stopping the reaction when an optimal contrast has been reached, spectrophotometric reading yields quantifiable results.
3. Prepare a standard curve from the data produced from the serial dilutions with concentrations on the *X*-axis vs. absorbance on the *Y*-axis. Interpolate the concentrations of the sample from this standard curve.
 - (a) Be consistent when adding standards to the assay plate. Add standards first and then samples to calibrate the results. Add standards to plate in the order from low concentration to high concentration as this will minimize the risk of compromising the standard curve.
 - (b) Results of ELISA can be measured both qualitatively and quantitatively; the qualitative results provide a simple positive or negative result in the form of a presence or absence of a visible colored product for a sample, whereas in quantitative ELISA, the optical density (or fluorescence) is measured and plotted into a standard curve, which is typically a serial dilution of the protein to be quantified.

4. Always prepare fresh stocks in limited quantity and discard any remaining solutions/samples. Always prepare fresh buffers at the correct pH. Samples should be properly stored and refrigerated at 2–4 °C for 1–2 days. For long period, preserve or store at –20 °C. Allow all the components to come to room temperature before use.
5. The detection systems are based on the utilization of flow cytometry, chemiluminescence, or electrochemiluminescence technology depending on the type of the label attached to the detecting antibody. The label can be multiple things such as fluorophors or enzymes. The most commonly used detections are Flow cytometric multiplex arrays, also known as bead-based multiplex assays, measured with chromogenic or fluorogenic emissions detected using flow cytometric analysis.
6. Antigens, serum samples, and assay chemicals used in ELISA may be potentially infectious and/or toxic. Wear gloves and lab coat before handling all serum and plasma specimens of blood-borne infections.

References

1. Engvall E, Perlman P (1971) Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry* 8:871–874
2. Van Weemen BK, Schuurs AH (1971) Immunoassay using antigen-enzyme conjugates. *FEBS Lett* 15:232–236
3. Towner KJ, Cockayne A (1993) Molecular methods for microbial identification and typing, 1st edn. Chapman & Hall, London, pp 1–202
4. van Belkum A, Tassios PT, Dijkhoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feil E, Gerner-Smidt P, Brisse S, Struelens M (2007) Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin Microbiol Infect* 13:1–46
5. Arbeit RD (1995) Laboratory procedures for the epidemiologic analysis of microorganisms. In: Murray PR, Baron EJ, Pfaller MA, Tenover FC, Tenover RH (eds) *Manual of clinical microbiology*, 6th edn. American Society of Microbiology, Washington, DC, pp 190–208
6. Lazcka O, Del Campo FJ, Munoz FX (2007) Pathogen detection: a perspective of traditional methods and biosensors. *Biosens Bioelectron* 22:1205–1217
7. Chen CH, Ding HC, Chang TC (2001) Rapid identification of *Bacillus cereus* based on the detection of a 28.5 kilodalton cell surface antigen. *J Food Prot* 64:348–354
8. Bolton FJ, Sails AD, Fox AJ, Wareing DRA, Greenway DLA (2002) Detection of *Campylobacter jejuni* and *Campylobacter coli* in foods by enrichment culture and polymerase chain reaction enzyme-linked immunosorbent assay. *J Food Prot* 65:760–767
9. Daly P, Collier T, Doyle S (2002) PCR-ELISA detection of *Escherichia coli* in milk. *Lett Appl Microbiol* 34:222–226
10. Peplow MO, Correa-Prisant M, Stebbins ME, Jones F, Davies P (1999) Sensitivity, specificity, and predictive values of three *Salmonella* rapid detection kits using fresh and frozen poultry environmental samples versus those of standard plating. *Appl Environ Microbiol* 65:1055–1065
11. Obst U, Hubner I, Wecker M, Bitter-Suermann D (1989) Immunological method using monoclonal antibodies to detect Enterobacteriaceae in drinking water. *Aqua* 38:136–142
12. Kfir R, Genthe B (1993) Advantages and disadvantages of the use of immunodetection techniques for the enumeration of microorganisms and toxins in water. *Water Sci Technol* 27:243–252

Chapter 14

Analysis of Microbial Diversity and Construction of Metagenomic Library

Thangamani Rajesh, Jeyaprakash Rajendhran, and
Paramasamy Gunasekaran

Abstract

Culture independent metagenomic methods represent promising approach for the identification of novel genes and genomes of uncultivable microorganisms. A wide range of protocols have been developed for the isolation of community DNA from environmental sources. Three important steps in DNA isolation from soil are (a) cell lysis, (b) removal of contaminants, in particular humic acid, and (c) purification of isolated community DNA. Depending on the nature of the environmental DNA source, modifications are usually performed in any of the above steps. The steps involved in the isolation of DNA and RNA from soil samples and the appropriate purification strategies are described in this chapter. Additionally, protocols for the microbial diversity analysis using 16S ribosomal RNA analysis and construction of metagenomic library for screening of functional genes from metagenomic DNA are given.

14.1 Introduction

Metagenomics studies in recent years had led to the discovery of novel enzymes, antibiotics, and uncultivable bacterial phylogenetic groups. A basic prerequisite in metagenomic approaches is the recovery of environmental DNA at optimal quantity and purity [2]. Extensive methods have been developed for the recovery of environmental DNA at sufficient quantity for further applications. Depending on the nature of the environment considered for investigation, technologies for purification of extracted environmental DNA have also been described [4, 5]. The protocols rely on the physical treatment, bead beating for cell lysis. The addition of glass beads of defined diameter and application of vigorous vibration cause rapid lysis of microbial cells in solution. As glass beads are of very small dimension, additional pressure is applied sufficiently to disrupt bacterial cells without causing any damage to the DNA. Following bead beating, the glass beads, together with soil and larger sized debris, are removed by

centrifugation. This allows the complete disruption of microbial cells, resulting in efficient lysis and recovery of total DNA.

Other protocols involve release of the tightly bound cells and spores by means of ultrasonication. Employing ultrasonic waves at defined frequency, microbial cells are disrupted to release DNA. Since the frequency of the ultrasonic waves could be controlled, this method ensures complete lysis of bacterial cells, fungal cells, and spores. This method also facilitates the release of tightly bound cells from the soil particles. Other techniques allow the isolation of DNA from cells previously separated from soil. Lysis of bacterial cells in molten agarose plugs ensures the recovery of high molecular weight metagenomic DNA. In this process, the bacterial cells are embedded in low melting agarose plugs and the cells are lysed within the agarose gel matrix. Since the lysis is performed within the plugs, high molecular weight DNA is obtained, which is suitable for metagenomic library construction using BAC vectors.

Co-purification of humic substances is a major problem in soil metagenomic DNA isolation. These contaminants are not completely removed during the DNA extraction protocols. Humic compounds interfere with downstream applications such as PCR and restriction digestion. Therefore, purification of the isolated DNA is essential for the desired applications. Recently, several simple and rapid purification methods have been reported for the successful removal of contaminants from metagenomic DNA [1, 3]. When sephadex beads are used for the purification, sephadex G-50 matrix exhibits differential binding properties to DNA and humic acid. Depending on the time of passage through this solid matrix and based on centrifugation speed, DNA could be recovered from contaminating substances. Other techniques involve the electroelution of DNA molecules separated in an agarose gel, which ensure high purity where the low melting agarose gels should be used. As the DNA molecules are entrapped in a solid matrix, physical force is required to release the DNA from the agarose gel. The application of electric field enables the release of DNA, which is called electroelution. The DNA sample is further subjected to dialysis to remove salts and concentrated either by precipitation or using vacuum concentrators [6, 7]. Metagenomic studies help to understand the genetic resource of uncultured microbes. Metatranscriptomics deals with the understanding of microbial gene expression pattern in environmental microbial communities. The metatranscriptomics studies involve the direct extraction of RNA (metagenomic RNA) and further downstream applications [8].

Available fingerprinting methods for microbial subtyping are ultimately based on the differences in the genome sequence. Therefore, DNA sequencing would appear to be the best approach to differentiate subtypes. Since it is impractical to

sequence the genome of every microbial strain, sequencing of stable marker genes is considered to be a potential strategy for microbial species identification. The rRNA is the highly conserved gene in all forms of life. Certain regions in the rDNA sequences are highly conserved even between distantly related organisms. This allows the precise positioning of distantly related organisms in the evolutionary tree and it also ensures the true measure of differences between closely related species. In bacteria, 16S rRNA sequences are used extensively to determine taxonomy and phylogeny (evolutionary relationships) and to estimate rates of species divergence among bacteria. It is the widely used molecular chronometer to measure the evolutionary relationship as bacterial identification and differentiation are generally based on the amplification of 16S rRNA gene sequences followed by sequencing and comparison of the sequence with known 16S sequences in databases.

PCR cloning kits take the advantage of the terminal transferase activity of *Taq* DNA polymerase. *Taq* DNA polymerase adds a single 3'-A overhang to both ends of the PCR product. The structure of these PCR products favors direct cloning into a linearized cloning vector with single 3'-dT overhang. Such an overhang at the vector cloning site not only facilitates cloning but also prevents the recircularization of the vector. As a result, more than 90 % of recombinant clones contain the vector with an insert. Recombinant clones are selected based on blue/white screening. By comparing the amplified 16S rRNA sequences, it is possible to estimate the historical branching order of the species, and also difference in the evolved sequences. The 16S rRNA-based phylogenetic tree can also be used to relate the three domains of life—bacteria, archaea, and eukarya. This provides information of lineage diverged from a common ancestral lineage. The lengths of the individual lines in the phylogenetic tree indicate the amount of sequence change.

Metagenomic DNA constitutes a promising source of novel functional genes. The function-driven and the sequence-driven analyses are the two major approaches to screen metagenomic libraries. The function-driven analysis is based on the identification of clones that express a desired trait, followed by characterization of the active clones by sequence and biochemical analysis. Sequence-driven analysis relies on the use of conserved DNA sequences to design hybridization probes or PCR primers to screen metagenomic libraries for clones that carry sequences of interest. Metagenomic libraries with larger DNA inserts such as bacterial artificial chromosome (BAC) libraries or cosmid libraries are useful for the sequence-based approaches. For functional approach, the number of functionally expressed genes should be as high as possible. Therefore, metagenomic libraries constructed using plasmid vectors such as pUC19 are widely used for

functional screening. Here, the metagenomic DNA library construction using pUC19 is provided. The recombinant clones may be screened for different genes by function-based or sequence-based approaches. Library construction with environmental DNA, heterologous expression in a suitable host, and screening for desired phenotype based on phenotypic acquisition constitutes basic steps in functional metagenomics. By principle, a large quantity of pure DNA is considered for library construction as it offers the analysis of a relatively large subset of an environmental population. The DNA (~10 µg or more) is fragmented using a suitable restriction enzyme. The vector DNA (plasmid, cosmid, or BAC) is also restricted with the same DNA and is followed by ligation reaction. Transformation is then performed with the ligated product and the recombinants are screened for desired phenotype. Alternatively, short linkers can be attached to the restricted fragments (in case of insert DNA <3 kb) and PCR is performed with primers targeting the linkers. This leads to the uniform amplification of a complex mixture of larger sized DNA fragments using small amounts of starting material. This is due to the ligation of small DNA fragments leading to the formation of long linear and circular DNA concatamers. As this process leads to accumulation of large amounts of DNA, restriction digestion and ligation of this DNA would be efficient for library construction.

14.2 Materials

14.2.1. Isolation of Metagenomic DNA from Soil

14.2.1.1. CTAB Method

1. Extraction buffer (50 mM Na-phosphate buffer [pH 8], 50 mM NaCl, 500 mM Tris-HCl [pH 8], 5 % sodium dodecyl sulfate)
2. Phenol-chloroform-isoamyl alcohol (25:24:1)
3. Sterile glass beads (0.25 mg of 0.1-mm diameter and 0.25 mg of 0.5-mm diameter)
4. Chloroform-isoamyl alcohol (24:1)
5. CTAB
6. Isopropanol
7. 70 % ethanol
8. TE buffer (10 mM Tris; 1 mM EDTA [pH 8.0])

14.2.1.2. TNEP Method

1. TNEP (50 mM Tris, 20 mM EDTA, 100 mM NaCl)
2. 1 % polyvinylpyrrolidone (w/v)

3. Lysozyme
4. Achromopeptidase
5. 1 % SDS
6. Isopropanol
7. 70 % ethanol
8. TE buffer (10 mM Tris; 1 mM EDTA [pH 8.0])

**14.2.1.3. Indirect Lysis
Method**

1. Chelex 100
2. 0.1 % nadeoxycholate
3. 2.5 % polyethylene glycol
4. Sterile gauze bandage
5. Lysis buffer (1 % sarkosyl, 1 % sodium deoxycholate, 1 mg/ml lysozyme, 10 mM Tris-HCl [pH 8.0], 0.2 M EDTA [pH 8.0], and 50 mM NaCl)
6. ESP buffer (1 % Sarkosyl, 1 mg/ml proteinase K, and 0.5 M EDTA [pH 8.0])
7. 1 mM phenylmethylsulfonyl in isopropanol
8. TE buffer (10 mM Tris-HCl with 50 mM EDTA [pH 8.0])
9. Isopropanol

**14.2.2. Purification of
Metagenomic DNA**

**14.2.2.1. Using Sephadex
G-50 Beads**

1. Sephadex G-50 beads
2. 10 mM potassium phosphate buffer (pH 7.2)
3. 10 mM potassium phosphate buffer
4. TE buffer (10 mM Tris-HCl with 50 mM EDTA [pH 8.0])

**14.2.2.2. Separation
in Agarose Gel**

1. Agarose (molecular biology grade)
2. Scalpel
3. Dialysis tubing
4. 1 mM EDTA
5. 2 % NaHCO₃
6. 1× TAE buffer
7. TE buffer (10 mM Tris-HCl with 50 mM EDTA [pH 8.0])

**14.2.3. Isolation
of Metagenomic
RNA from Soil**

1. 120 mM sodium phosphate buffer (pH 5.2)
2. 1 % diethyl pyrocarbonate
3. Denaturing solution (4 M guanidine thiocyanate, 25 mM sodium citrate, 0.5 % sarkosyl, 0.1 M 2-mercaptoethanol)
4. 2 M sodium acetate (pH 4.0)

5. Chloroform–isoamyl alcohol (24:1)
6. Isopropanol
7. DEPC-treated sterile water

**14.2.4. Amplification
of 16S rRNA
Sequences from
Metagenomic DNA**

1. Metagenomic DNA
2. 16S rDNA forward primer (8F: 5'-AGAGTTTGATCMT-GGCTCAG-3')
3. 16S rDNA reverse primer (1522R: 5'-AAGGAGGTGATG-GANCCRCA-3')
4. dNTPs
5. Taq DNA polymerase
6. PCR reaction buffer

**14.2.5. Cloning of
16S rRNA Amplicon**

1. PCR amplified and purified product
2. PCR cloning vector (pTZ57R/T, MBI Fermentas)
3. T4 DNA ligase
4. 5× ligation buffer
5. Sterile deionized water
6. Overnight culture of *Escherichia coli* DH5
7. CaCl₂ (100 mM/85 mM); MgCl₂ (100 mM)
8. LB medium
9. Sterile microcentrifuge tubes and tips
10. Sterile glycerol (15 %)
11. LB Agar with ampicillin (100 µg/ml), X-gal (20 µg/ml), and IPTG (40 µg/ml)

**14.2.6. Phylogenetic
Analysis of 16S rRNA
Sequences**

**14.2.7. Construction
of Metagenomic
Library**

1. 16S sequences of the analyte in FASTA format
1. Plasmid DNA (pUC18)
2. Metagenomic DNA
3. Restriction enzymes
4. Restriction enzyme buffers
5. T4 DNA ligase
6. T4 DNA ligase buffer
7. Electrocompetent *E. coli* DH10B cells
8. Ampicillin (100 µg/ml)
9. X-gal (20 µg/ml)
10. IPTG (40 µg/ml)
11. LB Agar

14.3 Method

14.3.1. Isolation of Metagenomic DNA from Soil by CTAB Method

1. To 0.6 g of soil, add 600 μ l of extraction buffer and 300 μ l of phenol–chloroform–isoamyl alcohol.
2. Add 0.5 g sterile glass beads to the tubes and homogenize the soil with the extraction buffer in a mini-bead beater for 90 s at 2,500 rpm.
3. Centrifuge the homogenate at $16,000 \times g$ for 2 min.
4. Collect the supernatant and mix with equal volume of phenol–chloroform–isoamyl alcohol and centrifuge the tubes at $6,000 \times g$ for 5 min.
5. Collect the supernatant and mix with an equal volume of chloroform–isoamyl alcohol and centrifuge at $16,000 \times g$ for 5 min.
6. Collect the supernatant and add NaCl to a final concentration of 1.5 M along with CTAB to 1 % and incubate the tubes at 65 °C for 30 min.
7. After the incubation, allow the solution to cool down to room temperature and mix the solution with an equal volume of chloroform–isoamyl alcohol (24:1), and centrifuge the tubes at $6,000 \times g$ for 20 min.
8. Collect the supernatant and add 0.6 volume of isopropanol and incubate at room temperature for 10 min. Centrifuge at $16,000 \times g$ for 10 min to pellet the precipitated DNA.
9. Discard the supernatant and wash the DNA pellet with 70 % ethanol and collect the DNA by centrifugation at $16,000 \times g$ for 5 min.
10. After the air drying to remove traces of ethanol, resuspend the isolated DNA in 200 μ l of TE buffer.

14.3.2. Isolation of Metagenomic DNA from Soil by TNEP Method

1. To 0.5 g of soil, add 0.5 ml of TNEP buffer and 1 % polyvinylpyrrolidone (w/v).
2. Sonicate the samples for 7–10 min at a power setting of 15 W with 50 % active cycles.
3. Followed by ultrasonication, add lysozyme and achromopeptidase to a final concentration of 0.3 mg/ml.
4. Incubate the tubes for 20 min at 37 °C and add SDS to a final concentration of 1 %.
5. Incubate the suspension for 1 h at 60 °C, vortex the tubes for 10 min, and centrifuge the suspension at $4,000 \times g$ for 5 min.
6. Collect the supernatant and add 0.6 volume of isopropanol to precipitate the DNA.

7. Wash the DNA with 70 % ethanol as described in the above-procedure and resuspend the DNA in a final volume of 100 μ l TE.

**14.3.3. Isolation
of Metagenomic
DNA from Soil
by Indirect Lysis
Method**

1. To 50 g of soil, add 10 g Chelex100, 100 ml of 0.1 % nadeoxycholate, and 2.5 % polyethylene glycol 6,000 in a 50 ml oak ridge tube.
2. Allow the samples to shake for 1 h at 100 rpm at 4 °C with occasional rapid agitation by hand.
3. Centrifuge the samples at $960 \times g$ for 15 min to pellet the soil particles.
4. Collect the supernatant and pass the supernatant through a sterile gauze bandage to remove Chelex100.
5. Harvest the bacterial fraction by centrifugation at $22,000 \times g$ for 20 min.
6. Embed the harvested bacterial cells in low-melting-point agarose in a 1-ml syringe.
7. Extrude the agarose plug from the syringe, add 10 ml of lysis buffer and 50 mM NaCl, and incubate the plugs for 1 h at 37 °C.
8. Transfer the plug to 40 ml of ESP buffer and incubate the plugs for 16 h at 55 °C.
9. After the incubation, inactivate proteinase K by adding 1 mM phenylmethylsulfonyl in isopropanol and incubate the plugs for 1 h at room temperature.
10. Wash the plugs three times with TE buffer for every 10 min and store the plugs at 4 °C in 10 mM Tris-HCl with 50 mM EDTA (pH 8.0).
11. Transfer the agarose plugs into a petri dish containing 10 ml TE and wash the plugs three times by gentle shaking.
12. After washing, add 5 ml of TE and 6 ml of isopropanol and incubate the plugs at 50 °C for 30 min.
13. Collect the precipitated DNA by centrifugation at $4,000 \times g$ for 5 min.
14. Wash the DNA with 70 % ethanol as described in the above procedure and resuspend the DNA in a final volume of 100 μ l TE.

**14.3.4. Purification
of Metagenomic
DNA by Using
Sephadex
G-50 Beads**

1. Wash and equilibrate 2 g of Sephadex G-50 beads with 10 ml of 10 mM potassium phosphate buffer (pH 7.2).
2. Resuspend 500 μ l of slurry in 1.5 ml of 10 mM potassium phosphate buffer.

3. After thorough mixing, transfer 50 μ l of the slurry in each 5 microfuge tubes and centrifuge for 1 min at $300 \times g$ to separate the overlaying buffer.
4. Add the metagenomic DNA preparations at different concentrations (100–500 μ g in 1 ml TE) to microfuge tubes containing Sephadex G-50 beads.
5. Mix the tubes by inverting the tubes for 15 min and allow the contents to pass through by incubating the tubes at room temperature for 5 min.
6. The brown color in the metagenomic DNA preparation could be seen bound to the matrix.
7. Collect the supernatant containing the DNA by centrifugation of the tubes at $1,000 \times g$ for 1 min.
8. The metagenomic DNA can then be analyzed for its purity by spectrophotometric analysis or restriction digestion analysis.

14.3.5. Purification of Metagenomic DNA by Separation in Agarose Gel

1. Resolve the Metagenomic DNA in 0.7 % agarose gel.
2. Excise the DNA fragment from the gel with a clean, sharp scalpel.
3. Preheat the dialysis tubing at 90 °C for 10 min in 1 mM EDTA/2 % NaHCO₃.
4. Rinse the tubes several times with sterile water prior to use.
5. Equilibrate ~300 mg of gel with 50 ml of fresh 1 \times TAE buffer at 4 °C for 30 min.
6. Remove the gel piece and place it lengthwise in a dialysis bag and add 400 μ l of 1 \times TAE.
7. Seal the dialysis bag and completely submerge it in an electrophoresis chamber.
8. Perform electrophoresis at field strength of ~4–5 V/cm.
9. After 2 h, reverse the polarity of the electrodes for 1 min to dissociate the DNA bound to the membrane.
10. Remove the dialysis bag from the electrophoresis chamber and dialyse against TE buffer for 2 h.
11. Collect the dialysed DNA and concentrate by vacuum concentrator.

14.3.6. Isolation of Metagenomic RNA from Soil

1. To 10 g of soil, add 20 ml of 120 mM sodium phosphate buffer (pH 5.2) and 1 % diethyl pyrocarbonate and allow the tubes for shaking at 150 rpm for 15 min.
2. Centrifuge the tubes at $6,000 \times g$ for 10 min. Discard the supernatant and wash the pellet once again with phosphate buffer.
3. To the pellet, add 15 ml of denaturing solution and allow the tubes for shaking at 200 rpm for 1 min.

4. To the mixture, add 1.5 ml of 2 M sodium acetate (pH 4.0) and mix the samples.
5. Extract the total RNA with 15 ml phenol and 3 ml of chloroform-isoamyl alcohol.
6. Vigorously shake the mixture to obtain a homogenous phase and leave the tubes containing the lysates on ice for 15 min.
7. Following incubation, centrifuge the tubes at $10,000 \times g$ for 20 min at 4 °C and extract the aqueous phase. To this aqueous phase, add equal volume of ice-cold isopropanol and store the tubes at -20 °C for 1 h to precipitate the total RNA.
8. Recover the RNA containing pellets by centrifugation at $10,000 \times g$ for 20 min at 4 °C followed by vacuum drying.
9. Resuspend the total RNA in 100 µl of DEPC-treated sterile water.

14.3.7. Amplification of 16S rRNA Sequences from Metagenomic DNA

1. Prepare a master mix with the following composition in the given order. A typical master mix for a 100 µl PCR is given below (Table 14.1). Modify the required volume proportionately.
2. Aliquot 18 µl of the master mix in each vial. Add 2 µl of the genomic DNA from different strains in each tube. Keep one negative control without template DNA. Perform PCR using the following cycling conditions.

Denaturation	94 °C for 5 min	
Denaturation	94 °C for 30 s	} 30 cycles
Annealing	55 °C for 30 s	
Extension	72 °C for 2 min	
Extension	72 °C for 5 min	
Holding	4 °C for 5 min	
End		

3. Add 2 µl of 6× loading dye to the 10 µl of PCR amplified product and mix.
4. Load each sample in a well in 1.0 % agarose gel and run the gel at 100 V for about 90 min.
5. Load 1 µl of 1 kb ladder as molecular weight marker to estimate the size of the amplified fragments.
6. Document the image in Gel documentation system.

Table 14.1
Composition of master mix for PCR

Ingredients	Volume (μ l)
10 \times buffer	10.0
10 mM dNTPs mixture	2.5
5 μ M 8F primer	10.0
5 μ M 1522R primer	10.0
Deionized water	56.5
Taq DNA polymerase (5 U/ μ l)	1.0
Total volume	90.0

Table 14.2
Composition of ligation reaction

Components	Volume (μ l)
Vector pTZ57R/T	2.0
5 \times Ligation buffer	6.0
Gel eluted PCR product	12.5
Deionized water	7.5
T4 DNA ligase (5 U/ μ l)	2.0
Total volume	30.0

14.3.8. Cloning of 16S rRNA Amplicon

14.3.8.1. Preparation of Competent Cells

Set up the ligation reaction as follows (Table 14.2) and incubate the ligation mix at 16 °C overnight.

1. Inoculate 0.5 ml of *E. coli* DH5 α overnight bacterial culture into 50 ml of LB medium (antibiotic free).
2. Grow up to early log phase at 37 °C in a shaker till the OD becomes 0.5–0.6 at 600 nm (2–3 h).
3. Centrifuge the culture in a sterile tube at 3,000 rpm for 5 min.
4. Discard the supernatant; add 9 ml of MgCl₂ (100 mM) and mix gently to suspend the cells completely in the solution.
5. Centrifuge at 3,000 rpm for 5 min at 4 °C.
6. Suspend the cell pellet in 9 ml of ice-cold CaCl₂ (100 mM) and incubate on ice for 40 min.

7. Centrifuge at 3,000 rpm for 5 min. Resuspend the cells in 3.5 ml of FT buffer (85 mM CaCl₂ and 15 % glycerol).
8. Make 100 µl aliquots of the cells. Freeze the cells in dry ice/ethanol or in liquid nitrogen.
9. Store the cells at –80 °C (these cells will retain their competence for 6–9 months).

14.3.8.2. Transformation

1. Place the aliquots of competent cells on ice and allow the cells to thaw for 5 min.
2. Add 10 µl of the ligation mix.
3. Mix gently and incubate on ice for 40 min.
4. Expose the cells for heat shock at 42 °C for 90 s and return to ice immediately.
5. Add 900 µl of antibiotic-free LB medium to each tube and incubate at 37 °C for 1 h with shaking.
6. Centrifuge the tubes at 3,000 rpm for 2 min and remove 800 µl of the supernatant.
7. Resuspend the pellet in the remaining 200 µl of medium.
8. Spread 100 µl of resuspended cells onto LB agar plates containing ampicillin, X-gal, and IPTG.
9. Incubate the plates in an inverted position at 37 °C overnight and screen the colonies after 10–12 h incubation.

14.3.9. Phylogenetic Analysis of 16S rRNA Sequences

Sequence analysis should be performed with positive recombinants. With the obtained sequence, further analysis can be performed as follows:

- Perform a BLAST analysis using the obtained sequence as query at the RDP database (<http://rdp.cme.msu.edu/>).
- Select the top ten hits that represent high query coverage and percentage identity and retrieve the sequences in FASTA format.
- Perform multiple sequence alignment with ClustalW program (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>).
- Phylogenetic assessment of the obtained sequence can then be made using the query sequence and other retrieved sequences employing Tree Builder software (<http://rdp.cme.msu.edu/treebuilder/treeing.spr>).

14.3.9.1. 16S rRNA Sequence

- An example of phylogenetic analysis is outlined below:
attgaacgctggcggcaggcctaacacatgcaagtcgagcggcagcacgggtactt
gtacctggtggcgagcggcgacgggtgagtaatgcctaggaatctgcctgtagt
gggggataacgttcgaaacggacgctaataccgcatacgtcctacgggagaaagc
aggggaccttcgggccttgcgctatcagatgagcctaggctcgattagctagttggt
gaggtaatggctaccaaggcgacgatccgtaactggctgagaggatgatcagtca

```

+ no rank Root (0/20/109270) (selected/match/total RDP sequences)
+ domain Bacteria (0/20/107507)
+ phylum "Proteobacteria" (0/20/49878)
+ class Gammaproteobacteria (0/20/27579)
+ order Pseudomonadales (0/20/7342)
+ family Pseudomonadaceae (0/20/5683)
+ genus Pseudomonas (0/20/5486)
  S000000053 not_calculated 1.000 1395 Pseudomonas amygdali (T); LMG 2123T (type strain); 276654
  S000381629 not_calculated 0.986 1433 Pseudomonas syringae pv. atropurpurea; MAFF 301017; AB001440
  S000381633 not_calculated 0.965 1433 Pseudomonas syringae pv. maculicola; MAFF 302264; AB001444
  S000381634 not_calculated 0.961 1435 Pseudomonas syringae pv. morsprunorum; MAFF 302280; AB001445
  S000381639 not_calculated 0.961 1434 Pseudomonas syringae pv. theae; PT1; AB001450
  S000498035 not_calculated 0.966 1413 Pseudomonas syringae pv. tomato str. DC3000; AE016853
  S000498037 not_calculated 0.966 1413 Pseudomonas syringae pv. tomato str. DC3000; AE016853
  S000498039 not_calculated 0.966 1413 Pseudomonas syringae pv. tomato str. DC3000; AE016853
  S000498042 not_calculated 0.966 1413 Pseudomonas syringae pv. tomato str. DC3000; AE016853
  S000498044 not_calculated 0.966 1413 Pseudomonas syringae pv. tomato str. DC3000; AE016853
  S000640019 not_calculated 0.964 1345 Pseudomonas sp. Enf4; DQ339631
  S001611507 not_calculated 0.972 1385 Pseudomonas syringae pv. porri; P55; FN554248
  S002155707 not_calculated 0.967 1247 Pseudomonas cannabina pv. alisalensis; B591; GQ470207
  S002155708 not_calculated 0.967 1247 Pseudomonas cannabina pv. alisalensis; B5413; GQ470208
  S002155709 not_calculated 0.962 1247 Pseudomonas cannabina pv. alisalensis; B51034; GQ470209
  S002155710 not_calculated 0.966 1247 Pseudomonas syringae pv. maculicola; B5286; GQ470210
  S002155712 not_calculated 0.967 1247 Pseudomonas cannabina pv. alisalensis; B5130; GQ470212
  S002155713 not_calculated 0.962 1247 Pseudomonas cannabina pv. alisalensis; B5416; GQ470213
  S002155714 not_calculated 0.967 1247 Pseudomonas syringae pv. tomato; B5287; GQ470214
  S002155715 not_calculated 0.962 1247 Pseudomonas cannabina pv. alisalensis; B5215; GQ470215

```

Fig. 14.1 Computed output of RDP database.

```

cactggaactgagacacgggtccagactcctacgggaggcagcagtggggaatattg
gacaatgggagaaagcctgatccagccatgccgcgtgtgtgaagaaggtcttcgga
ttgtaaacgactttaagttgggaggaagggcattacctaatacgttaagtgtttgacg
ttaccgacagaataagcaccggctaactctgtgccagcagccgcggttaatacagagg
gtgcaagcgtaatacgaattactgggcgtaaacgcgcgtaggtggtttgttaagt
tgaatgtgaaatccccgggctcaactgggaactgcataaaaactggcaggctaga
gtatggtagaggggtggtggaatttctgtctagcggtagaatgcgtagatataggaa
ggaacaccagtggcgaaggcgaccactggactgatactgacactgaggtgcgaaa
gcgtggggagcaaacaggattagataccctggtagtcacgcccgtaaacgatgtcaa
ctagccgttgggagccttgagctcttagtggcgagctaacgcattaagttgaccgcc
tggggagtacggccgaagggttaaaactcaaatgaattacggggggcccgcaag
cgggtggagcatgtggtttaattcgaagcaacgcgaagaaccttaccaggccttgacat
ccaatgaatccttagagatagaggagtgccctcgggagcattgagacaggtgctgc
atggctgtcgtcagctcgtgtcgtgagatgttggttaagtcccgtaacgagcgcaa
ccctgtccttagttaccagcacgttaaggtgggactctaaggagactgccggtgac
aaaccggaggaaggtggggatgacgtcaagtcacatggcccttacggcctgggct
acacacgtgtacaatggtcggtagagaggttgccaaccgcgaggtggagctaat
ctcaaaaaccgatcgtagtcggatcgagctctgaactcgactgcgtgaatcgga
atcgctagtaatcggaatcagaatgtcgcggtgaatacgttccgggctgtgtacac
accgccgtcacaccatgggagtggttgaccagaagtagctagttaaccttcgg
cgggacggttaccacggtgtgattcatgactggggtgaagtcgtaacaaggtagcc
gtaggggaacctgc

```

- Submit this sequence to RDP database. The computed output will be as shown below (Fig. 14.1):
- Select the top ten hits and retrieve the sequences in FASTA format.

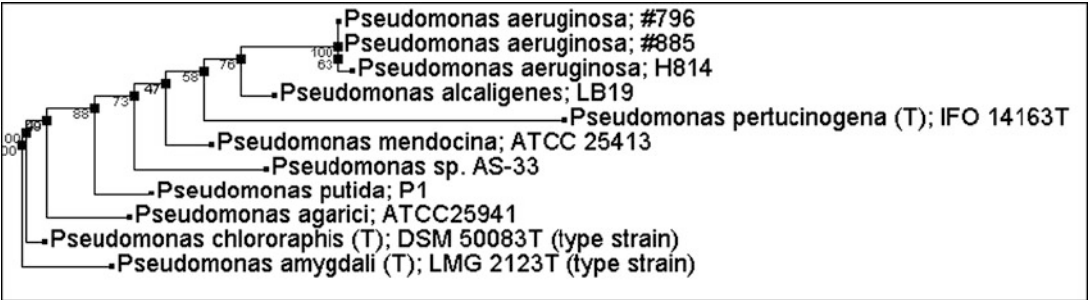


Fig. 14.2 Phylogenetic tree construction with Tree Builder.

Table 14.3
Composition of restriction digestion mixture

Components	Volume (μl)
pUC19 plasmid	10.0
10× buffer	2.0
BamHI (10 U/μl)	1.0
Deionized water	7.0
Total volume	20.0

- ClustalW multiple sequence alignment with these sequences will yield results as shown below:
- Employing these sequences, phylogenetic tree construction with Tree Builder software, the results will be obtained as shown below (Fig. 14.2):

**14.3.10. Construction
of Metagenomic
Library**

**14.3.10.1. Restriction
Digestion of Vector**

1. Digest 5 μg of pUC19 plasmid with BamHI. Set the restriction digestion reaction on ice. A typical 20 μl reaction mix is given below (Table 14.3):
2. Incubate the reaction mixture at 37 °C for 2 h and inactivate reaction by heating at 70 °C for 10 min.
3. Dephosphorylate the linearized vector (to prevent self-ligation) by adding 0.5 U of calf intestine alkaline phosphatase and incubate at 37 °C for 1 h.
4. Purify the restriction digested and dephosphorylated vector by phenol/chloroform extraction or gel purification.

Table 14.4
Composition of restriction digestion of metagenomic DNA

Components	Volume (μl)
Metagenomic DNA	10.0
10× buffer	2.0
BamHI (10 U/μl)	1.0
Deionized water	7.0
Total volume	20.0

Table 14.5
Composition of ligation mixture

Components	Volume (μl)
Dephosphorylated vector	2.0
5× Ligation buffer	2.5
3–8 kb metagenomic fragments	12.5
Deionized water	11.0
T4 DNA ligase	2.0
Total volume	30.0

14.3.10.2. Restriction Digestion of Metagenomic DNA

1. The isolated metagenomic DNA should be digested with the same restriction enzyme.
2. Digest 5 μg of metagenomic DNA with BamHI. Set the restriction digestion reaction on ice. A typical 20 μl reaction mix is given below (Table 14.4):
3. Incubate the reaction mixture at 37 °C for 2 h and inactivate the reaction by heating at 70 °C for 10 min.
4. Resolve the digested DNA fragments on a 0.7 % agarose gel. Excise the fragments of about 3–8 kb and extract using QIAquick Gel purification kit.

14.3.10.3. Ligation

1. Set up the ligation reaction as follows and incubate the ligation mixture (Table 14.5) at 16 °C overnight.
2. Inactivate the ligation by heating at 70 °C for 10 min and use the ligation mixture for the electrotransformation of *E. coli* DH10B cells.

S000000053	-----ATTGAACGCTGGCGGCAGGCCTAACACATGCAA	33
S000000900	-----ATTGAACGCTGGCGGCAGGCCTAACACATGCAA	33
S000001173	-----TAACACATGCAA	12
S000000079	-----TTTGATCCTGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAA	49
S000000446	-----TTTGATCATGG-TCAGATTGAACGCTGGCGGCAGGCATAACACATGCAA	48
S000000348	-----AGTTTGATCCTGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAA	51
S000000739	-----AGTTTGATCCTGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAA	51
S000000939	-----AGTTTGATCCTGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAA	51
S000000875	-----ATTGAACGCTGGCGGCAGGCCTAACACATGCAA	33
S000000866	-----ATTGAACGCTGGCGGCAGGCCTAACACATGCAA	33
S000001098	-----ATTGAACGCTGGCGGCAGGCCTAACACATGCAA	33

S000000053	GTCGAGCGGCAGCACGGGTACTTGTACCTGGTGGCG-AGCGGCGGACGGGTGAGTAATGC	92
S000000900	GTCGAGCGGTAG-AGAGGTGCTTCCACCTCTTGAG--AGCGGCGGACGGGTGAGTAATGC	90
S000001173	GTCGAGCGGATG-AAGGGAGCTTGCTCCCGGATTG--AGCGGCGGACGGGTGAGTAATGC	69
S000000079	GTCGAGCGGATG-AGAGGAGCTTGCTCCTTGATTT--AGCGGCGGACGGGTGAGTAATGC	106
S000000446	GTCGAGCGGATG-AAGAGAGCTTGCTCTCTGATTG--AGCGGCGGACGGGTGAGTAATGC	105
S000000348	GTCGAGCGGATG-AAGGGAGCTTGCTCCTGGATTG--AGCGGCGGACGGGTGAGTAATGC	108
S000000739	GTCGAGCGGATG-AAGGGAGCTTGCTCCTGGATTG--AGCGGCGGACGGGTGAGTAATGC	108
S000000939	GTCGAGCGGATG-AAGGGAGCTTGCTCCTGGATTG--AGCGGCGGACGGGTGAGTAATGC	108
S000000875	GTCGAGCGGATG-AGTGGAGCTTGCTCCATGATTG--AGCGGCGGACGGGTGAGTAATGC	90
S000000866	GTCGAGCGGATG-ACGGGAGCTTGCTCCTTGATTG--AGCGGCGGACGGGTGAGTAATGC	90
S000001098	GTCGAGCGGAAG-AAGGGAGCTTGCTCCCGGATTG--AGCGGCGGACGGGTGAGTAATGC	90
***** * * * * *		
S000000053	CTAGGA-ATCTGCCTGGTAGTGGGGGATAACGTTTCGAAACGGACGCTAATACCGCATAC	151
S000000900	CTAGGA-ATCTGCCTGGTAGTGGGGGATAACGTTTCGAAACGGACGCTAATACCGCATAC	149
S000001173	CTAGGA-ATCTGCCTGGTAGTGGGGGACAACGTTTCGAAAGGAACGCTAATACCGCATAC	128
S000000079	CTAGGA-ATCTGCCTGGTAGTGGGGGATAACGTTCCGAAAGGAACGCTAATACCGCATAC	165
S000000446	CTAGGA-ATCTGCCTGATAGTGGGGGACAACGTTTCGAAAGGAACGCTAATACCGCATAC	164
S000000348	CTAGGA-ATCTGCCTGGTAGTGGGGGATAACGTTCCGAAACGGGCGCTAATACCGCATAC	167
S000000739	CTAGGA-ATCTGCCTGGTAGTGGGGGATAACGTTCCGAAACGGGCGCTAATACCGCATAC	167
S000000939	CTAGGA-ATCTGCCTGGTAGTGGGGGATAACGTTCCGAAACGGGCGCTAATACCGCATAC	167
S000000875	CTAGGA-ATCTGCCTGGTAGTGGGGGACAACGTTTCGAAAGGAACGCTAATACCGCATAC	149
S000000866	CTAGGA-ATCTGCCTGGTAGTGGGGGACAACGTTTCGAAAGGAACGCTAATACCGCATAC	149
S000001098	CTGGA-ATCTGCCTGGTAGTGGGGGATNNNGTCCGAAACGGGNNNTAATACCGCGTAC	149
** ** * ***** ** * * ***** **		
S000000053	GTCCTACGGGAGAAAGCAGGGGACCTTCGGGCCTTGCGCTATCAGATGAGCCTAGGTCGG	211
S000000900	GTCCTACGGGAGAAAGCAGGGGACCTTCGGGCCTTGCGCTATCAGATGAGCCTAGGTCGG	209
S000001173	GTCCTACGGGAGAAAGCAGGGGACCTTCGGGCCTTGCGCTATCAGATGAGCCTAGGTCGG	188
S000000079	GTCCTACGGGAGAAAGCAGGGGACCTTCGGGCCTTGCGCTATCAGATGAGCCTAGGTCGG	225
S000000446	GTCCTACGGGAGAAAGCAGGGGACCTTCGGGCCTTGCGCTATCAGATGAGCCTAGGTCGG	224
S000000348	GTCCTGAGGGAGAAAGTGGGGGATCTTCGGACCTCACGCTATCAGATGAGCCTAGGTCGG	227
S000000739	GTCCTGAGGGAGAAAGTGGGGGATCTTCGGACCTCACGCTATCAGATGAGCCTAGGTCGG	227
S000000939	GTCCTGAGGGAGAAAGTGGGGGATCTTCGGACCTCACGCTATCAGATGAGCCTAGGTCGG	227
S000000875	GTCCTACGGGAGAAAGCAGGGGACCTTCGGGCCTTGCGCTATCAGATGAGCCTAGGTCGG	209
S000000866	GTCCTACGGGAGAAAGCAGGGGACCTTCGGGCCTTGCGCTATCAGATGAGCCTAGGTCGG	209
S000001098	GTCCTACGGGAGAAAGCAGGGGATCTTCGGACCTTGCGCTATCAGATGAGCCGAGGCCG	209
***** ***** ***** ***** ** ***** ***** ** *		
S000000053	ATTAGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCCGTAACCTGGTCTGAGAGG	271
S000000900	ATTAGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCCGTAACCTGGTCTGAGAGG	269
S000001173	ATTAGCTAGTTGGTGAGGTAAGGCTCACCAAGGCGACGATCCGTAACCTGGTCTGAGAGG	248
S000000079	ATTAGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCCGTAACCTGGTCTGAGAGG	285
S000000446	ATTAGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCCGTAACCTGGTCTGAGAGG	284
S000000348	ATTAGCTAGTTGGTGAGGTAAGGCTACCAAGGCGACGATCCGTAACCTGGTCTGAGAGG	287
S000000739	ATTAGCTAGTTGGTGAGGTAAGGCTACCAAGGCGACGATCCGTAACCTGGTCTGAGAGG	287
S000000939	ATTAGCTAGTTGGTGAGGTAAGGCTACCAAGGCGACGATCCGTAACCTGGTCTGAGAGG	287
S000000875	ATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATCCGTAACCTGGTCTGAGAGG	269
S000000866	ATTAGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCCGTAACCTGGTCTGAGAGG	269
S000001098	ATTAGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCCGTAACCTGGTCTGAGAGG	269
***** ***** ***** ** ***** ***** ***** *****		

```

S000000053 ATGATCAGTCACACTGGAAC TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG 331
S000000900 ATGATCAGTCACACTGGAAC TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG 329
S000001173 ATGATCAGTCACACTGGAAC TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG 308
S000000079 ATGATCAGTCACACTGGAAC TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG 345
S000000446 ATGATCAGTTACACTGGAAC TGAGACACGGTCCAGACTCGTATGGGAGGCAGCAGTGGGG 344
S000000348 ATGATCAGTCACACTGGAAC TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG 347
S000000739 ATGATCAGTCACACTGGAAC TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG 347
S000000939 ATGATCAGTCACACTGGAAC TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG 347
S000000875 ATGATCAGTCACACTGGAAC TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG 329
S000000866 ATGATCAGTCACACTGGAAC TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG 329
S000001098 ATGATCAGTCACACTGGAAC TGAGACACGGTCCANACTCCTACGGGANGCAGCAGTGGGG 329
*****
S000000053 AATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTC 391
S000000900 AATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTC 389
S000001173 AATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTC 368
S000000079 AATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTC 405
S000000446 AATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTC 404
S000000348 AATATTGGACAANGGGCGAAANNNTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTC 407
S000000739 AATATTGGACAATGGGCGAAANNNTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTC 407
S000000939 AATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTC 407
S000000875 AATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTC 389
S000000866 AATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTC 389
S000001098 AATATTGGACAATGGGGGAAACCCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTC 389
*****
S000000053 GGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCATTACCTAATACGTAAGTGTTT -TG 450
S000000900 GGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGTACTTACCTAATACGTAGTATTTT -TG 448
S000001173 GGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCATTACCTAATACGTTAGTGTTT -TG 427
S000000079 GGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTT -TG 464
S000000446 GGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCATTACCTAATACGTTAGTGTTT -TG 463
S000000348 GGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTT -TG 466
S000000739 GGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTT -TG 466
S000000939 GGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTT -TG 466
S000000875 GGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTT -TG 448
S000000866 GGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTT -TG 448
S000001098 GGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCTTGCGGCTAATACCTCGCAAGTTT 449
*****
S000000053 ACGTTACCGACAGAATAAGCACCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAGA GG 510
S000000900 ACGTTACCGACAGAATAAGCACCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAGAGG 508
S000001173 ACGTTACCGACAGAATAAGCACCGGCTAACTCTGTNCCAGCAGCCGCGGTAATACAGAGG 487
S000000079 ACGTTACCGACAGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGG 524
S000000446 ACGTTACCGACAGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGG 523
S000000348 ACGTTACCAACAGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGG 526
S000000739 ACGTTACCAACAGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGG 526
S000000939 ACGTTACCAACAGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGG 526
S000000875 ACGTTACCGACAGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGG 508
S000000866 ACGTTACCGACAGAATAAGCACCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAGAGG 508
S000001098 ACGTTACCAACAGAATAAGCACCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAGAGG 509
*****
S000000053 GTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGA 570
S000000900 GTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTCGTTAAGTTGG 568
S000001173 GTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTGGTTAAGTTGG 547
S000000079 GTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTCGTTAAGTTGG 584
S000000446 GTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGA 583
S000000348 GTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTCAGCAAGTTGG 586
S000000739 GTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTCAGCAAGTTGG 586
S000000939 GTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTCAGCAAGTTGG 586
S000000875 GTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTCAGCAAGTTGG 568
S000000866 GTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGG 568
S000001098 GTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTCAGTAAGATGG 569
*****

```



```

S000000053    CAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTA 927
S000000900    CAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTA 925
S000001173    CAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTA 904
S000000079    CAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTA 944
S000000446    CAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTA 940
S000000348    CAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTA 943
S000000739    CAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTA 943
S000000939    CAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTA 943
S000000875    CAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCA TGTGGTTTA 925
S000000866    CAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTA 925
S000001098    CAAGATTAAAACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTA 926
*****

S000000053    ATTCGAAGCAACGCGAAGAACCTTACCAGGCCTTGACATCCAATGAATCCTTTAGAGATA 987
S000000900    ATTCGAAGCAACGCGAAGAACCTTACCAGGCCTTGACATCCAATGAATCTTCCAGAGATG 985
S000001173    ATTCGAAGCAACGCGAAGAACCTTACCAGGCCTTGACATCCAATGAATCTTCCAGAGATG 964
S000000079    ATTCGAAGCAACGCGAAGAACCTTACCTGGCCTTGACATGCTGAGAACTTTCAGAGATG 1004
S000000446    ATTCGAAGCAACGCGAAGAACCTTACCAGGCCTTGACATGCAGAGAACTTTCAGAGATG 1000
S000000348    ATTCGAAGCAACGCGAAGAACCTTACCTGGCCTTGACATGCTGAGAACTTTCAGAGATG 1003
S000000739    ATTCGAAGCAACGCGAAGAACCTTACCTGGCCTTGACATGCTGAGAACTTTCAGAGATG 1003
S000000939    ATTCGAAGCAACGCGAAGAACCTTACCTGGCCTTGACATGCTGAGAACTTTCAGAGATG 1003
S000000875    ATTCGAAGCAACGCGAAGAACCTTACCTGGCCTTGACATGCTGAGAACTTTCAGAGATG 985
S000000866    ATTCGAAGCAACGCGAAGAACCTTACCAGGCCTTGACATGCAGAGAACTTTCAGAGATG 985
S000001098    ATTCGAAGCAACGCGAAGAACCTTACCTGGCCTTGACATGCAGAGAACCTCCAGAGATG 986
*****

S000000053    GAGGAGTGCCTTCGGGAGCATTGAGACAGGTGCTGCGATGGCTGTCGTGAGCTCGTGTCTG 1047
S000000900    GATTGGTGCCTTCGGGAACATTGAGACAGGTGCTGCGATGGCTGTCGTGAGCTCGTGTCTG 1045
S000001173    GAGGAGTGCCTTCGGGAACATTGAGACAGGTGCTGCGATGGCTGTCGTGAGCTCGTGTCTG 1024
S000000079    GATTGGTGCCTTCGGGAACCTCAGACACAGGTGCTGCGATGGCTGTCGTCAGCTCGTGTCTG 1064
S000000446    GATTGGTGCCTTCGGGAACCTCAGACACAGGTGCTGCGATGGCTGTCGTGAGCTCGTGTCTG 1060
S000000348    GATTGGTGCCTTCGGGAACCTCAGACACAGGTGCTGCGATGGCTGTCGTGAGCTCGTGTCTG 1063
S000000739    GATTGGTGCCTTCGGGAACCTCAGACACAGGTGCTGCGATGGCTGTCGTGAGCTCGTGTCTG 1063
S000000939    GATTGGTGCCTTCGGGAACCTCAGACACAGGTGCTGCGATGGCTGTCGTGAGCTCGTGTCTG 1063
S000000875    GATTGGTGCCTTCGGGAACCTCAGACACAGGTGCTGCGATGGCTGTCGTGAGCTCGTGTCTG 1045
S000000866    GATTGGTGCCTTCGGGAACCTCAGACACAGGTGCTGCGATGGCTGTCGTGAGCTCGTGTCTG 1045
S000001098    GATGGGTGCCTTCGGGAACCTCAGACACAGGTGCTGCGATGGCTGTCGTGAGCTCGTGTCTG 1046
**      *****

S000000053    GAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCTTAGTTACCAGCACGTTA 1107
S000000900    GAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCTTAGTTACCAGCACGTTA 1105
S000001173    GAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCTTAGTTACCAGCACGTTA 1084
S000000079    GAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCTTAGTTACCAGCACGTTA 1124
S000000446    GAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCTTAGTTACCAGCACGTTA 1120
S000000348    GAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCTTAGTTACCAGCACCTCG 1123
S000000739    GAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCTTAGTTACCAGCACCTCG 1123
S000000939    GAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCTTAGTTACCAGCACCTCG 1123
S000000875    GAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCTTAGTTACCAGCACGTTA 1105
S000000866    GAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCTTAGTTACCAGCACGTTA 1105
S000001098    GAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCTTAGTTACCAGCACCTCG 1106
*****

S000000053    AGGTGGGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAA 1167
S000000900    TGGTGGGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAA 1165
S000001173    TGGTGGGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAA 1144
S000000079    TGGTGGGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAA 1184
S000000446    AGGTGGGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAA 1180
S000000348    GG-TGGGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAA 1182
S000000739    GG-TGGGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAA 1182
S000000939    GG-TGGGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAA 1182
S000000875    TGGTGGGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAA 1165
S000000866    TGGTGGGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAA 1165
S000001098    GG-TGGGCACTCTAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAA 1165
*      *****

```

S000000053	GTCATCATGGCCCTTACGGCCTGGGCTACACACGTGCTACAATGGTCGGTACAGAGGGTT	1227
S000000900	GTCATCATGGCCCTTACGGCCTGGGCTACACACGTGCTACAATGGTCGGTACAGAGGGTT	1225
S000001173	GTCATCATGGCCCTTACGGCCTGGGCTACACACGTGCTACAATGGTCGGTACAAAGGGTT	1204
S000000079	GTCATCATGGCCCTTACGGCCAGGGCTACACACGTGCTACAATGGTCGGTACAAAGGGTT	1244
S000000446	GTCATCATGGCCCTTACGGCCTGGGCTACACACGTGCTACAATGGTCGGTACAAAGGGTT	1240
S000000348	GTCATCATGGCCCTTACGGCCAGGGCTACACACGTGCTACAATGGTCGGTACAAAGGGTT	1242
S000000739	GTCATCATGGCCCTTACGGCCAGGGCTACACACGTGCTACAATGGTCGGTACAAAGGGTT	1242
S000000939	GTCATCATGGCCCTTACGGCCAGGGCTACACACGTGCTACAATGGTCGGTACAAAGGGTT	1242
S000000875	GTCATCATGGCCCTTACGGCCAGGGCTACACACGTGCTACAATGGTCGGTACAAAGGGTT	1225
S000000866	GTCATCATGGCCCTTACGGCCTGGGCTACACACGTGCTACAATGGTCGGTACAGAGGGTT	1225
S000001098	GTCATCATGGCCCTTACGGCCAGGGCTACACACGTGCTACAATGGGGGATACAAAGGGTT	1225

S000000053	GCCAAACCGCGAGGTGGAGCTAATCTCAGAAAACCGATCGTAGTCCGGATCGCAGTCTGC	1287
S000000900	GCCAAAGCCGCGAGGTGGAGCTAATCCCAAAAACCGATCGTAGTCCGGATCGCAGTCTGC	1285
S000001173	GCCAAACCGCGAGGTGGAGCTAATCCCAAAAACCGATCGTAGTCCGGATCGCAGTCTGC	1264
S000000079	GCCAAAGCCGCGAGGTGGAGCTAATCCCAAAAACCGATCGTAGTCCGGATCGCAGTCTGC	1304
S000000446	GCCAAAGCCGCGAGGTGGAGCTAATCCCAAAAACCGATCGTAGTCCGGATCGCAGTCTGC	1300
S000000348	GCCAAAGCCGCGAGGTGGAGCTAATCCCAAAAACCGATCGTAGTCCGGATCGCAGTCTGC	1302
S000000739	GCCAAAGCCGCGAGGTGGAGCTAATCCCAAAAACCGATCGTAGTCCGGATCGCAGTCTGC	1302
S000000939	GCCAAAGCCGCGAGGTGGAGCTAATCCCAAAAACCGATCGTAGTCCGGATCGCAGTCTGC	1302
S000000875	GCCAAAGCCGCGAGGTGGAGCTAATCCCAAAAACCGATCGTAGTCCGGATCGCAGTCTGC	1285
S000000866	GCCAAAGCCGCGAGGTGGAGCTAATCTCAGAAAACCGATCGTAGTCCGGATCGCAGTCTGC	1285
S000001098	GCCAAAGCCGCGAGGTGGAGCTAATCCCAAAAGTCTCTCGTAGTCCGGATTGGAGTCTGC	1285

S000000053	AACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATA	1347
S000000900	AACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATA	1345
S000001173	AACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATA	1324
S000000079	AACTCGACTGCGTGAAGTCGGAATCGCTGGTAATCGTGAATCAGAATGTCACGGTGAATA	1364
S000000446	AACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGTGAATCAGAATGTCACGGTGAATA	1360
S000000348	AACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGTGAATCAGAATGTCACGGTGAATA	1362
S000000739	AACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGTGAATCAGAATGTCACGGTGAATA	1362
S000000939	AACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGTGAATCAGAATGTCACGGTGAATA	1362
S000000875	AACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGTGAATCAGAATGTCACGGTGAATA	1345
S000000866	AACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATA	1345
S000001098	AACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAACGCCACGGTGAATA	1345

S000000053	CGTTCCTCCGGGCTTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCACCAGAAGTA	1407
S000000900	CGTTCCTCCGGGCTTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCACCAGAAGTA	1405
S000001173	CGTTCCTCCGGGCTTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCACCAGAAGTA	1384
S000000079	CGTTCCCTCCCTTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCTCCAGAAGTA	1371
S000000446	CGTTCCTCCGGGCTTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCTCCAGAAGTA	1420
S000000348	CGTTCCTCCGGGCTTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCTCCAGAAGTA	1422
S000000739	CGTTCCTCCGGGCTTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCTCCAGAAGTA	1422
S000000939	CGTTCCTCCGGGCTTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCTCCAGAAGTA	1422
S000000875	CGTTCCTCCGGGCTTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCTCCAGAAGTA	1405
S000000866	CGTTCCTCCGGGCTTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCACCAGAAGTA	1405
S000001098	CGTTCCTCCGGGCTTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCACCAGAAGTA	1405

14.3.10.4. Preparation of Electrocompetent DH10B Cells

1. Pick a colony of DH10B and inoculate into a 5 ml of LB broth. Grow overnight at 37 °C with shaking.
2. Next morning, add 1 % of the overnight culture into 500 ml of LB medium and incubate at 37 °C with shaking until OD 600 reaches ~0.7 (this takes ~2 h).
3. Cool cells in cold room on ice for ~20 min.

4. Centrifuge the rotor without tubes for 5 min to precool to 4 °C.
5. Pour cells into 2 precooled centrifuge bottles, 250 ml each. Spin at 5,000 rpm for 15 min at 4 °C.
6. Decant supernatant and resuspend cells in 500 ml of sterile ice-cold water.
7. Spin at 5,000 rpm for 15 min at 4 °C. Decant supernatant and resuspend cells in 250 ml of sterile ice-cold water.
8. Centrifuge at 5,000 rpm for 15 min at 4 °C. Decant supernatant and resuspend cells in 125 ml of sterile ice-cold water.
9. Centrifuge at 5,000 rpm for 15 min at 4 °C. Decant and resuspend cells in 10 ml ice-cold 10 % glycerol. Transfer cells into a 50 ml centrifuge tube and centrifuge on tabletop centrifuge, at 13,000 rpm for 15 min at 4 °C. Decant supernatant and resuspend cells in 0.5–1 ml sterile 10 % glycerol.
10. Aliquot 100 µl of cells into each precooled microcentrifuge tubes on ice and store electrocompetent cells at –80 °C until use.

14.3.10.5.

Electrotransformation

1. Prior to electroporation, ligation mixture must be precipitated with ethanol or diluted to prevent the samples from causing an arc to jump across the cuvette upon application of the pulse.
2. Thaw an aliquot of *E. coli* DH10B cells on ice. When cells are thawed, add 1–10 µl of ligation mixture to the cells and mix by tapping gently.
3. Carefully pipette the cell/DNA mixture into a chilled 0.1 cm cuvette. Gently tap the cuvette to ensure that the cell/DNA mixture makes contact all the way across the bottom of the cuvette chamber. Avoid formation of bubbles.
4. Wipe the outside of the cuvette with a tissue to dry it, place it in the electroporation chamber, and apply pulse. For BioRad GenePulser® II electroporator, the recommended pulse conditions are 2.0 kV, 200 Ω, and 25 µF.
5. Immediately after pulsing, add 900 µl of SOC medium and transfer the solution to a microcentrifuge tube. Delaying this transfer can seriously reduce the survival of transformed cells.
6. Incubate at 225 rpm (37 °C) for 1 h with shaking.
7. Spread the cells on LB agar plates containing ampicillin (100 µg/ml), X-gal (20 µg/ml), and IPTG (40 µg/ml).
8. Incubate plates overnight at 37 °C, to permit the color to develop sufficiently to distinguish blue colonies from white.

References

1. Cullen DW, Hirsch PR (1998) Simple and rapid method for direct extraction of microbial DNA from soil for PCR. *Soil Biol Biochem* 30: 983–993
2. Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68:669–685
3. Liles MR, Williamson LL, Rodbumrer J, Torsvik V, Goodman RM, Handelsman J (2008) Recovery, purification and cloning of high molecular weight DNA from soil microorganisms. *Appl Environ Microbiol* 74:3302–3305
4. Rajendhran J, Gunasekaran P (2008) Strategies for accessing soil metagenome for desired applications. *Biotechnol Adv* 6:576–590
5. Rajesh T, Rajendhran J, Lavanya Pushpam P, Gunasekaran P (2011) Methods in metagenomic DNA and RNA isolation from soil. In: Frans. J. de Bruijn (ed) *Handbook of molecular microbial ecology II: metagenomics in different habitats*, John Wiley & Sons Inc. vol 2, pp 93–107
6. Sagova-Mareckova M, Cermak L, Novotna J, Plhachova K, Forstova J, Kopecky J (2008) Innovative methods for soil DNA purification tested in soils with widely differing characteristics. *Appl Environ Microbiol* 74: 2902–2907
7. Sharma PK, Capalash N, Kaur J (2007) An improved method for single step purification of metagenomic DNA. *Mol Biotechnol* 36: 61–63
8. Tsai Y, Park MJ, Olson BH (1991) Rapid method for direct extraction of mRNA from seeded soils. *Appl Environ Microbiol* 57: 765–768

Chapter 15

Bioinformatics Tools for Interpretation of Data Used in Molecular Identification

Suchi Smita, Krishna P. Singh, Bashir A. Akhoon, Shishir K. Gupta, and Shailendra K. Gupta

Abstract

The advancement in the high-throughput instrumentation techniques for genome sequences resulted in massive amount of data, which are publically available in biological databases. Availability of these large numbers of completely sequenced prokaryotic and eukaryotic genomes changed the way the biological and biomedical experiments were conducted in the past. Varieties of tools were generated to help researchers to extract the information and messages encoded in these sequences. The current chapter highlights some of the basic bioinformatics protocols to use various tools and utilities for molecular data interpretation.

15.1 Introduction

The amount of sequence data generated through high-throughput techniques is significantly outstripping the storage capacity available; hence it is the need to store the data in a particular format that require less storage space, be easily understandable, and quickly accessible. Biological sequences are encoded in specific formats based on the sequencing methodology involved; also various utilities/applications understand different sequence formats. Thus, for successful job submission, it is important to understand the variety of sequence formats used for describing biological sequences and the procedure to change one sequence format into another using various sequence format converting utilities.

Sequence format is the representation in which nucleotide and protein sequences are stored in the computers. All sequence formats are standard ASCII files that differ in the way they hold sequence data and other related information about sequence, such as sequence ID, organism name, sequence title, date of submission, function, and comments. Some of the frequently

```
>gi|3372530|gb|AF052295.1| Bradyrhizobium japonicum ferric uptake
regulator (fur) gene, complete cds
GGGCCCATCGCGGCCGCGGCCTCTCCACACGCTGCTGATGAGCCATCTCGGCCACCTCGCCGGGCGCGG
CGTGCGCACGATATTTCTCGAGGTTCGAGGAAAACAACCAGCCGGCACGGCGGCTCTACGCGAGGTGCGGA
TTCATGGTGGTCGGCCGCGCGAACGCTACTATAAACAGCCGAACGGGGAACAATTGAACGCCCTTCTGA
TGCGGCGTGACTTGTCGTAACATTGATGGCAGAAAGCGCCCCGTGAGGCAGACAGATCATGACCGCACTG
AAACCTTCTTCTGCATCCAAGGCGTCCGGCATCGAGGCGCGCTGTGCCGCCACCGGCATGCGCATGACCG
AGCAGCGCCGCGTCATCGCCCGCGTGCTCGCGGAGGCGGTTCGATCATCCCGACGTCGAGGAATTATACCG
CCGTGCGTCGCGTCGACGACAAGATCTCGATCTCAACCGTGTATCGCACCGTCAAGCTGTTTCGAGGAT
GCCGGCATCATCGAACGCCATGACTTTCGCGAGGGACGCGCGCGCTACGAGACGATGCGCGACAGCCATC
ACGACCACCTCATCAATCTGCGCGACGGCAAGGTGATCGAGTTCACCTCCGAAGAGATCGAGAAGCTCCA
GGCGGAGATCGCCCGCAAGCTCGGCTACAAGCTGGTCGACCACCGGCTCGAGCTCTATTGCGTCCCGCTC
GACGACGACAAGCCACAAAGCTAAGTGCCCGTCGATCTCATCATCTTCGACTGCGATGGCGTGTCTCGTG
ACAGCGAGGTGATCTCCTGTGTCGCGCATGCGGATGTGCTGACCCGCCACGGCTATCCGATC
```

Fig. 15.1 Sample FASTA file containing ferric uptake regulator gene from *Nostoc azollae*. Note the “>” symbol in the beginning of entry, followed by title of the sequence. Each line of the sequence contains <80 characters.

encountered sequence formats while handling various bioinformatics tools and software are:

**15.1.1. FASTA
(Pearson) Format**

This is the most general format to represent DNA, RNA, and protein sequences. Sequence is written in standard IUPAC single-letter codes for nucleotide and protein sequences. The first line of each sequence begins with a greater-than sign (>), followed by single line header containing sequence information. Rest of the lines contains sequence itself. It is recommended to store <80 characters per line. Depending on the application, blank lines in FASTA files can be ignored or considered as sequence termination. Also, the spaces and nonsequence symbols are either ignored or treated as gaps as shown in Fig. 15.1. FASTA files may contain multiple sequences with one sequence listed right after another. The multi-FASTA format is accepted by most of the multiple sequence alignment tools.

**15.1.2. GenBank
Format**

GenBank is an annotated collection of all publically available DNA sequences managed by National Institute of Health, MD, USA. GenBank file provides information related to gene and gene product (e.g., sequence, length, definition, keywords, source organism, references, etc.). GenBank format (GenBank Flat File Format) consists of an annotation section and a sequence section. The start of the annotation section is marked by a line beginning with the word “LOCUS.” The start of sequence section is marked by a line beginning with the word “ORIGIN” and the end of the section is marked by a line with only “//” as shown in Fig. 15.2. Some of the commonly used fields in the GenBank records are summarized in Table 15.1.

```

LOCUS      AF052295                832 bp    DNA        linear    BCT 20-SEP-1999
DEFINITION Bradyrhizobium japonicum ferric uptake regulator (fur) gene,
            complete cds.
ACCESSION  AF052295
VERSION    AF052295.1  GI:3372530
KEYWORDS   .
SOURCE     Bradyrhizobium japonicum
  ORGANISM Bradyrhizobium japonicum
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE  1 (bases 1 to 832)
  AUTHORS  Hamza,I., Chauhan,S., Hassett,R. and O'Brian,M.R.
  TITLE    The bacterial irr protein is required for coordination of heme
            biosynthesis with iron availability
  JOURNAL  J. Biol. Chem. 273 (34), 21669-21674 (1998)
  PUBMED   9705301
REFERENCE  2 (bases 1 to 832)
  AUTHORS  Hamza,I., Hassett,R. and O'Brian,M.R.
  TITLE    Identification of a functional fur gene in Bradyrhizobium japonicum
  JOURNAL  J. Bacteriol. 181 (18), 5843-5846 (1999)
  PUBMED   10482529
REFERENCE  3 (bases 1 to 832)
  AUTHORS  O'Brian,M.R.
  TITLE    Direct Submission
  JOURNAL  Submitted (06-MAR-1998) Biochemistry, State University of New York
            at Buffalo, 140 Farber Hall, 3435 Main Street, Buffalo, NY 14214,
            USA
FEATURES             Location/Qualifiers
     source            1..832
                       /organism="Bradyrhizobium japonicum"
                       /mol_type="genomic DNA"
                       /strain="strain I110"
                       /db_xref="taxon:375"
     gene              269..724
                       /gene="fur"
     CDS                269..724
                       /gene="fur"
                       /function="transcriptional regulator"
                       /note="FUR"
                       /codon_start=1
                       /transl_table=11
                       /product="ferric uptake regulator"
                       /protein_id="AAC32180.1"
                       /db_xref="GI:3372531"
                       /translation="MTALKPSSASKASGIEARCAATGMRMTEQRRVIARVLAEAVDHP
DVEELYRRCVAVDDKISISTVYRTVKLFEDAGIIERHDFREGRRARYETMRDSSHHDHLI
NLRDGKVIETFTSEETIEKLQAEIARKLGYKLVDRHRELYCVPLDDDKPTS"
ORIGIN
1  gggcccatcg  cggccgcggc  ctctcccaca  cgctgctgat  gagccatctc  ggccacctcg
61  ccgggcgcgg  cgtgcgcacg  atatttctcg  aggtcgagga  aaacaaccag  ccggcacggc
121  ggctctacgc  gaggtgcgga  ttcatggtgg  tcggcgcccg  cgaacgctac  tataaacagc
181  cgaacgggga  acaattgaac  gcccttctga  tcgggcgtga  ctgtcgtaa  cattgatggc
241  agaaagcgcc  cgctcaggca  gacagatcat  gaccgcactg  aaaccttctt  ctgcattcaa
301  ggcgtccggc  atcgaggcgc  gctgtgcgcg  caccggcatg  cgcataccg  agcagcgccg
361  cgtcatcgcc  cgctgctcgc  cggaggcgg  cgatcatccc  gacgtcgagg  aattataccg
421  ccgtcgcgtc  gccgtcgacg  acaagatctc  gatctcaacc  gtgtatcgca  ccgtcaagct
481  gttcgaggat  gccggcatca  tcgaacgcca  tgactttcgc  gagggacgcg  cgcgctacga
541  gacgatcgcg  gacagccatc  acgaccacct  catcaatctg  cgcgacggca  aggtgatcga
601  gttcacctcc  gaagagatcg  agaagctcca  ggcgagatc  gccgcgaagc  tcggctacaa
661  gctggtcgac  caccggctcg  agctctattg  cgtcccgcctc  gacgacgaca  agcccacaag
721  ctaagtcccc  gtcgatctca  tcattctcga  ctgcgatggc  gtgctcgtgg  acagcgagg
781  gatctcctgt  cgcgcgcgat  cggatgtgct  gaccgcgcc  ggctatccga  tc
//

```

Fig. 15.2 Sample GenBank file format containing ferric uptake regulator gene from *Nostoc azollae*. Note the data available in locus and feature section of the file.

Table 15.1
A summary of fields commonly found in GenBank record

Field	Description
LOCUS	Contain unique locus name, often the first letter of genus and species followed by accession number; sequence length; type of the sequence (e.g., Genomic DNA, genomic RNA, mRNA, rRNA, tRNA, cytoplasmic RNA, etc.); Molecular topology (linear/ circular); GenBank division and date the file was last revised
DEFINITION	Brief description of the sequence including organism name and the gene/protein name
ACCESSION	Unique sequence identifier
VERSION	Version of the entry. Allow users to track multiple incarnation of a given sequence
KEYWORDS	Various keywords describing the sequence
SOURCE	Organism name followed by detail classification from NCBI taxonomy database
REFERENCES	Publications by the authors for the GenBank entry
FEATURES	A concise summary of the gene/protein annotation along with various biologically important regions in the sequence
ORIGIN	Beginning of the sequence data

15.1.3. EMBL Format

In EMBL format, each entry in the database is composed of lines represented by specific identifier to record various types of data, which make up the entry. All entries begin with the line identifier and end with a terminator line // as shown in Fig. 15.3. Some of the commonly used line identifiers in the EMBL format and their detailed description are listed in Table 15.2.

15.1.4. MSF File Format

Multiple sequence files (MSF) are used in various software programs for sequence analysis package. As with PHYLIP program for phylogenetic analysis, the use of multiple sequence alignment data is required as input in the MSF format. A sample file is shown in Fig. 15.4 where ten fur proteins are aligned. The file begins with either PileUP, !!NA_MULTIPLE_ALIGNMENT, or !!AA_MULTIPLE_ALIGNMENT. The second line begins with MSF: <Length of the alignment> (here it is 191 amino acids), followed by type of alignment (N: Nucleotide, P: Protein), a checksum number, and two dots that indicate the end of the header. Next block contains information of sequences, their name, length, a checksum, and weight.

15.1.5. PHYLIP File Format

The Phylogeny Inference Package (PHYLIP) is widely used to infer phylogenies, generating evolutionary trees and distance matrices for both nucleotide and protein sequences. All the

```

ID AF052295 standard; linear DNA; BCT; 832 BP.
XX
DT 20-SEP-1999
XX
DE Bradyrhizobium japonicum ferric uptake regulator (fur) gene,
DE complete cds.
XX
AC AF052295;
XX
SV AF052295.1 GI:3372530
XX
KW .
XX
OS Bradyrhizobium japonicum
OC Bradyrhizobium japonicum
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
XX
RN [1]
RP 1-832
RA Hamza,I., Chauhan,S., Hassett,R. and OBrian,M.R.
RT "The bacterial irr protein is required for coordination of heme
RT biosynthesis with iron availability";
RL J. Biol. Chem. 273 (34), 21669-21674 (1998)
XX
RN [2]
RP 1-832
RA Hamza,I., Hassett,R. and OBrian,M.R.
RT "Identification of a functional fur gene in Bradyrhizobium japonicum";
RL J. Bacteriol. 181 (18), 5843-5846 (1999)
XX
RN [3]
RP 1-832
RA OBrian,M.R.
RT "Direct Submission";
RL Submitted (06-MAR-1998) Biochemistry, State University of New York
RL at Buffalo, 140 Farber Hall, 3435 Main Street, Buffalo, NY 14214,
RL USA
XX
FH Key Location/Qualifiers
FT source 1..832
FT /organism="Bradyrhizobium japonicum"
FT /mol_type="genomic DNA"
FT /strain="strain I110"
FT /db_xref="taxon:375"
FT gene 269..724
FT /gene="fur"
FT CDS 269..724
FT /gene="fur"
FT /function="transcriptional regulator"
FT /note="FUR"
FT /codon_start=1
FT /transl_table=11
FT /product="ferric uptake regulator"
FT /protein_id="AAC32180.1"
FT /db_xref="GI:3372531"
FT /translation="MTALKPSSASKASGIEARCAATGMRMTEQRRVIARVLAEAVDHPD
FT VEELYRRCAVAVDDKISISTVYRTVKLFEDAGIIERHDFREGRARYETMRDSSHDLINL
FT RDGKVIEFTSEEIEKLQAEIARKLGYKLVDRHRELYCVPLDDDKPTS"
SQ Sequence 832 BP;
gggcccacgc cgcccgccgc ctctcccaca cgctgctgat gagccatctc ggccacctcg 60
ccggcgccgc cgtgcgcacg atatttctcg aggtcgagga aaacaaccag ccggcacggc 120
ggctctacgc gaggtcgcca ttcatggttg tcggccgccg cgaacgctac tataaacagc 180
cgaacgggga acaattgaac gcccttctga tcgcgcggtga cttgtcgtaa cattgatggc 240
agaaagcgcc cgtgcaggca gacagatcat gaccgcactg aaaccttctt ctgcatccaa 300
ggcgctccgc atcgaggcgc gctgtgccgc caccggcatg cgcgtgacgc agcagcgccg 360
cgtcatcgcc cgctgctgcg cggaggcggt cgatcatccc gacgtcgagg aattataccg 420
ccgctgcgtc gccgtcgacg acaagatctc gatctcaacc gtgtatcgca ccgtcaagct 480
gttcgaggat gcggcatcca tcgaacgcca tgactttcgc gaggggacgc cgcgctacga 540
gacgatcgcg gacagccatc acgaccacct catcaatctg cgcgacggca aggtgatcga 600
gttcacctcc gaagagatcg agaagctcca ggccgagatc gcccgcaagc tcggctacaa 660
gtcgtgcgac caccggctcg agctctattg cgtcccgcgc gacgacgaca agccccacaag 720
ctaagtgcgc gtcgatctca tcattcttca ctgcatggc gtgctcgtgg acagcgaggt 780
gatctcctgt cgcgccgatg cggatgtgct gacccgccac ggctatcga tc 832
//

```

Fig. 15.3 Sample EMBL file format containing ferric uptake regulator gene from *Nostoc azollae*. Note the line identifier before each line.

Table 15.2
A summary of fields commonly found in EMBL record

Line identifier	Description
ID	Identification line. Always the first line of any EMBL entry which include sequence id, sequence version, topology, type of molecule, dataclass, division, and sequence length
XX	No data or comments
AC	Accession number
DT	Date of data submission and last modification
DE	Data description contains general information about the sequence stored
SV	Sequence version information along with the gi (genInfo identifier) number
KW	Keyword line that provides information to generate cross-reference indexes of the sequence based on structural and functional criteria
OS	Organism species from which the sequence is derived
OC	Organism taxonomic classification
RN	Reference number, a unique number to each reference within the entry
RP	Reference position, a optional line type, which appears if one or more contiguous base spans of the presented sequence are listed in the reference
RA	Reference author name
RL	Reference location line contains information about journal, year of publication, volume, and page number
RT	Reference title
RX	Reference cross reference
DR	Database cross reference
CC	Comments about the entry
FH	Feature header
FT	Feature table contains summary of structural and functional regions within the sequence and provides cross linkages to other databases
SQ	Sequence header contains brief information about the sequence followed by sequence data in 5'–3'. Each line of the sequence data is composed of 60 bases grouped into a block of ten bases. Each block in the line is separated by space
//	Terminator of the entry

programs available in the PHYLIP package require sequences to be formatted in PHYLIP's own format. A sample PHYLIP file format is shown in Fig. 15.5. The file format is almost similar to the MSF file format. The first line of the file contains two numbers, i.e., number of sequences and the length of the alignment

PileUp

MSF: 191 Type: P Check: 2949 ..

Name: gi|295687474|ref|YP_003591167. oo Len: 191 Check: 9112 Weight: 9.0
 Name: gi|325059641|gb|ADY63332.1| oo Len: 191 Check: 6574 Weight: 9.0
 Name: gi|27375908|ref|NP_767437.1| oo Len: 191 Check: 9039 Weight: 10.0
 Name: gi|51588753|emb|CAH20364.1| oo Len: 191 Check: 6442 Weight: 6.0
 Name: gi|115348324|emb|CAL21256.1| oo Len: 191 Check: 6330 Weight: 6.0
 Name: gi|53720551|ref|YP_109537.1| oo Len: 191 Check: 1545 Weight: 11.0
 Name: gi|40388402|gb|AAR85472.1| oo Len: 191 Check: 4326 Weight: 10.0
 Name: gi|48479086|gb|AAT44865.1| oo Len: 191 Check: 1301 Weight: 16.0
 Name: gi|271962470|ref|YP_003336666. oo Len: 191 Check: 8379 Weight: 18.0
 Name: gi|298489843|ref|YP_003720020. oo Len: 191 Check: 9901 Weight: 15.0

//

```
gi|295687474|ref|YP_003591167. ....MDRLEKA CIEKGMRMTD QRRVIARVLS SA..EDHPDV
gi|325059641|gb|ADY63332.1| .....M IDLSKTLEEL CAERGMRMTD QRRVIARVLQ ES..ADHPDV
gi|27375908|ref|NP_767437.1| ..MTALKPSS ASKASGIEAR CAATGMRMTE QRRVIARVLA EA..VDHPDV
gi|51588753|emb|CAH20364.1| .....MTDNNKA LKNAGLKVTL PRLKILEVLQ NPA.CHHVSA
gi|115348324|emb|CAL21256.1| .....MTDNNKA LKNAGLKVTL PRLKILEVLQ NPA.CHHVSA
gi|53720551|ref|YP_109537.1| .....MTNPTD LKNIGLKATL PRLKILEIFQ QSP.VRHLLTA
gi|40388402|gb|AAR85472.1| .....MIDERMNSDE LKRAGLKATL PRLKILRIFE DSD.ARHLTA
gi|48479086|gb|AAT44865.1| .....MS AYTASSLKAE LNARGWRLTP QREKILHVFO NLPKGNHLTA
gi|271962470|ref|YP_003336666. ....MTETSWEHQ LRARGYRVTP QRQLVLEAVK AA...EHATP
gi|298489843|ref|YP_003720020. MQKPAISTKA ISSLEDALHR CQMLGMRVSR QRRFILELLW QAN..EHLTA
```

```
gi|295687474|ref|YP_003591167. EELHRRHAHI DPHISIATVY RTVRLFEESG IIERHDFRDG RSRYEETP..
gi|325059641|gb|ADY63332.1| EELYRRSSAV DPRISISTVY RTVKLFEDAG IIERHDFRDG RSRYETVP..
gi|27375908|ref|NP_767437.1| EELYRRCVAV DDKISISTVY RTVKLFEDAG IIERHDFREG RARYETMR..
gi|51588753|emb|CAH20364.1| EDLYKKLIDI GEEIGLATVY RVLNQFDDAG IVTRHNFEGG KSVFELTQ..
gi|115348324|emb|CAL21256.1| EDLYKILIDI GEEIGLATVY RVLNQFDDAG IVTRHNFEGG KSVFELTQ..
gi|53720551|ref|YP_109537.1| EDVYRNLLHE ELDIGLATVY RVLTKQFEQAG LLRSRNFESG KAVFELNE..
.
..
```

Fig. 15.4 Sample MSF file format containing ten ferric uptake regulator proteins aligned from different microorganisms.

followed by the sequence data. The sequence name is limited to ten characters only.

15.1.6. ALN/ClustalW2 Format

ALN/ClustalW2 file format is generated during multiple sequence alignment by ClustalW/ClustalX software. The format is widely accepted by various software suits analyzing multiple sequences in order to investigate structural and functional relationships. The ALN/ClustalW2 file begins with the software version of Clustal through which the file is generated followed by the aligned sequences in a block of 60 residues or nucleotides per line. Information about the residues alignment is written in the last line by three special characters. The character “*” indicates that all the nucleotides/residues are identical in all the sequences in the alignment; “:” means conserved substitutions have been observed, while “.” represents semi-conserved substitutions at that site. A sample file is shown in Fig. 15.6.

Often, the bioinformatics analysis on a given set of data requires integration of third party independent software. Most


```

10      191
gi|2956874 ----- ---MDRLEKA CIEKGRMTD QRRVIARVLS SA--EDHPDV
gi|3250596 -----M IDLSKTLEEL CAERGMRMTD QRRVIARVLQ ES--ADHPDV
gi|2737590 --MTALKPSS ASKASGIEAR CAATGMRMTE QRRVIARVLA EA--VDHPDV
gi|5158875 -----MTDNNKA LKNAGLKVTL PRLKILEVLQ NPA-CHHVSA
gi|1153483 -----MTDNNKA LKNAGLKVTL PRLKILEVLQ NPA-CHHVSA
gi|5372055 -----MTNPTD LKNIGLKATL PRLKILEIFQ QSP-VRHLTA
gi|4038840 -----MIDERMSDE LKRAGLKATL PRLKILRIFE DSD-ARHLTA
gi|4847908 -----MS AYTASSLKAE LNARGWRLTP QREKILHVFO NLPKGNHLSA
gi|2719624 -----MTETSWHEQ LRARGYRVTQ QRQLVLEAVK AA---EHATP
gi|2984898 MQKPAISTKA ISSLEDALHR CQMLGMRVSR QRRFILELLW QAN--EHLISA

EELHRRHAHI DPHISIATVY RTVRLFEEESG IIERHDFRDG RSRYEETP--
EELYRRSSAV DPRISISTVY RTVKLFEDAG IIERHDFRDG RSRYETVP--
EELYRRCVAV DDKISISTVY RTVKLFEDAG IIERHDFREG RARYETMR--
EDLYKKLIDI GEEIGLATVY RVLNQFDDAG IVTRHNFEGG KSVFELTQ--
EDLYKILIDI GEEIGLATVY RVLNQFDDAG IVTRHNFEGG KSVFELTQ--
EDVYRNLHE ELDIGLATVY RVLTFEQAG LLSRSNFESG KAVFELNE--
GEIYRLLLET GEEVGLATVY RVLTFEMAG LVRRHHFEGD KAVFELNE--
EELQELLDKR GEGISLSTIY RSVKLMSRMG ILRELELAEG HKHYELNQPY
EEICARVRET ARGVNISTVY RTLELLEELG MVTHTHLGHG APTYHLAA--
REIYDRNLQQ GKEIGHTSVY QNLEALSSQG IIECIEHCDG RLYGNIND--

THHHDHLIDM KTGKVVEFVD EEIEALQNAI ARKLGKYLVD HRLELYGVPL
EEHHDHLIDL KNSVVFIEFHS PEIEALQEKI AREHGFKLVD HRLELYGVPL
DSHHDHLINL RDGKVIEFTH EEIEKLQAEI ARKLGKYLVD HRLELYCVPL
QHHDHLICL DCGKVIEFSN ESIESLQREI AKQHGKILTN HSLYLYGHCE
QHHDHLICL DCGKVIEFSN ESIESLQREI AKQHGKILTN HSLYLYGHCE
GSHHDHLVCL DCGRVVEFFD AEIETRQOSI AKERGFKLQE HSLAMYGTCT
TGHHDHMVCT ACGKVLEFFD EMLEARQREL AANRGFFISD HSLYLYGTCL
PHHHHHLVCI QCNKTIENFN DSILKHSCLK CEKEGFQLID CQLTVMAICP
DSHDVHLVCH ECGEINEARP EVVQEFVTKL DEELGFAIDV HHLTVFGRCR
--AHSHVNCV DTNQLDVIH ELPAELIQVQ EAQTGVKIIA YTINFFGHRN

EE-----
KPG----- -EH-----
DDD----- -KPTS-----
TGN-----C REDESAHSKR
TGN-----C REDESAHSKR
TEN-----CPY RKH-----
GMQDVGICSL RDDAPGAST D-----
EALRMGWPSG IPSNWGCTRS LVDTRFQNC E IPESKEPEPE N
NCR-----
S-----

```

Fig. 15.5 Sample PHYLIP file format with aligned ferric uptake regulator proteins from ten different microorganisms.

of the time, the output generated by one program cannot be used directly as the input to another program. The format of the file need to be converted before second program can actually read the input file. There are several utilities available on the Web that automatically converts one file format into another. Some of these utilities are listed in Table 15.3.

15.2 Materials

15.2.1. Converting Sequence Format Using SeqVerter

1. Hardware: Computer system with at least Pentium 450 MHz processor, 50 MB hard disk space, 512 MN RAM.
2. Software: Window 7, Vista, XP, 200, Linux, Mac OS. SeqVerter sequence format conversion utility

CLUSTAL X (1.81) multiple sequence alignment

```

gi|295687474|ref|YP_003591167.      -----MDRLEKACIEKGMRMTDQRRVIARVLSSA--EDHPDV
gi|325059641|gb|ADY63332.1|         -----MIDLSKTEELCAERGMRMTDQRRVIARVLQES--ADHPDV
gi|27375908|ref|NP_767437.1|        --MTALKPSSASKASGIEARCAATGMRMTQRRVIARVLAE--VDHPDV
gi|51588753|emb|CAH20364.1|         -----MTDNNKALKNAGLVKVTLPRLKILEVLQNPA--CHHVSA
gi|115348324|emb|CAL21256.1|        -----MTDNNKALKNAGLVKVTLPRLKILEVLQNPA--CHHVSA
gi|53720551|ref|YP_109537.1|        -----MTNPTDLKNIGLKATLPRLKILEIFQQSP--VRHLTA
gi|40388402|gb|AAR85472.1|         -----MIDERMNSDELKRALGLKATLPRLKILRIFEDSD--ARHLTA
gi|48479086|gb|AAT44865.1|         -----MSAYTASSLKAELNARGWRLTPQREKILHVQNLPKGNHLSA
gi|271962470|ref|YP_003336666.      -----MTETSWHQQLRARGYKVTTPQRLVLEAVKAA---EHATP
gi|298489843|ref|YP_003720020.      MQKPAISTKAISSELDALHRCQMLGMRVSRQRRFILELLWQAN--EHLAS
                                     * : : * : . . *

gi|295687474|ref|YP_003591167.      EELHRRRAHAIDPHISIATVYRTVRLFEESGIERHDFDRGSRYEETP--
gi|325059641|gb|ADY63332.1|         EELYRRSSAVDPRI SISTVYRTVKLFEDAGIERHDFDRGSRYETVP--
gi|27375908|ref|NP_767437.1|        EELYRRCAVDDKISISTVYRTVKLFEDAGIERHDFREGRARYETMR--
gi|51588753|emb|CAH20364.1|         EDLYKKLIDIGEEIGLATVYRVLNQFDDAGIVTRHNFEGGKSVFELTQ--
gi|115348324|emb|CAL21256.1|        EDLYKILIDIGEEIGLATVYRVLNQFDDAGIVTRHNFEGGKSVFELTQ--
gi|53720551|ref|YP_109537.1|        EDVYRNLLHEELDIGLATVYRVLTQFQAGLRSNFESGKAVFELNE--
gi|40388402|gb|AAR85472.1|         GEIYRLLLETGEEVGLATVYRVLTQFEMAGLVRHHFEGDGKAVFELNE--
gi|48479086|gb|AAT44865.1|         EELQELLDKRGEGISLSTIYRSVKLMSRMGILRELELAEGHKHYELNQPY
gi|271962470|ref|YP_003336666.      EEICARVRETARGVNISTVYRTLELLEELGMVTHLHGHCAPTYYHLA--
gi|298489843|ref|YP_003720020.      REIYDRLNQQKGKEIGHTSVYQNLEALSSQGIIECIEHCDGRLYGNIND--
                                     : : . : : * : . . * : .

gi|295687474|ref|YP_003591167.      THHHDHLIDMKTGKVVEFVDEEIEALQNAIARKLYKGLVDHRLLEYGVPL
gi|325059641|gb|ADY63332.1|         EEHHDHLIDLKNSVVEIEFHSPEIEALQEIAAREHGFKLVDRHRLLEYGVPL
gi|27375908|ref|NP_767437.1|        DSHHDHLINLRDGKVIETSEEIEKLQAEIARKLYKGLVDHRLLEYGVPL
gi|51588753|emb|CAH20364.1|         QHHHDHLICLDCKGVIEFVSNEISLQREIAKQHGKLTNHSLYLYGHCE
gi|115348324|emb|CAL21256.1|        QHHHDHLICLDCKGVIEFVSNEISLQREIAKQHGKLTNHSLYLYGHCE
gi|53720551|ref|YP_109537.1|        GSHHDHLVCLDCGRVEEFFDAEIEIETRQQSIAKERGFKLQEHSLAMYGTCT
gi|40388402|gb|AAR85472.1|         TGHHDHVMVCTACGKVLEFFDEMLEARQRELAANRGFFISDHSLYLYGTCL
gi|48479086|gb|AAT44865.1|         PHHHHHLVCIQCCKTIEFNND SILKHSLKQCEKEGFLQDCQLTVMATCP
gi|271962470|ref|YP_003336666.      DSDHVHLVCHCEGGEINARPEVVQEFVTKDEELGFAIDVHHLTVFGRCR
gi|298489843|ref|YP_003720020.      --AHSHVNCVDTNQILDVHIELPAELIQQVEAQTGVKIATYTNFFGHRN
                                     * * : . : : * : : .

gi|295687474|ref|YP_003591167.      EE-----
gi|325059641|gb|ADY63332.1|         KPG-----EH-----
gi|27375908|ref|NP_767437.1|        DDD-----KPTS-----
gi|51588753|emb|CAH20364.1|         TGN-----CREDESAHSKR-----
gi|115348324|emb|CAL21256.1|        TGN-----CREDESAHSKR-----
gi|53720551|ref|YP_109537.1|        TEN-----CPYRKH-----
gi|40388402|gb|AAR85472.1|         GMQDVGICSLRDDAPAGASTD-----
gi|48479086|gb|AAT44865.1|         EALRMGWPSGIPSNWGTSLVDTRFQNCILPESKEPEPEN
gi|271962470|ref|YP_003336666.      NCR-----
gi|298489843|ref|YP_003720020.      S-----

```

Fig. 15.6 Sample ALN/ClustalW file format with aligned ferric uptake regulator proteins from ten different microorganisms generated using Clustal X version 1.81 program. Note the *last line* in the alignment file with three special characters, i.e., *asterisk*, *colon*, *dot* along with the conservation observed at that site.

[Software can be downloaded from: http://www.genestudio.com/download_seq.htm].

- Files: Download the *Bradyrhizobium japonicum* ferric uptake regulator (fur) gene (Accession number: AF052295) from GenBank database available at <http://www.ncbi.nlm.nih.gov> in GenBank format.

15.2.2. Raw Sequence Editing Tool

FinchTV (<http://www.geospiza.com/finchtv.html>) editing tool.

Table 15.3
Some commonly used sequence file format converting tools

Tool	Description	Web address
ReadSeq	The multi-format molbio sequence reader is a Web-based utility that automatically identify input single or multiple nucleotide/protein sequences format and convert into user defined format	http://searchlauncher.bcm.tmc.edu/seq-util/Options/readseq.html
Sequence Format Converter	A simple tool that convert between common sequence format. The tool is based on SeqIO module of BioPerl	http://www.bioinformaticsbox.com/tools/sequence_format_converter.php
Squizz	The Web-resource provides two separate utilities: (1) sequence format converter tool and (2) alignment format converter tool	http://mobyline.pasteur.fr/cgi-bin/portal.py?#forms::squizz_convert
SeqVerter	Standalone sequence file format conversion utility by GeneStudio, Inc. The software for various sequence editing functions	http://www.genestudio.com/seqverter.htm

15.2.3. Sequence Analysis Tools for DNA and Proteins

1. Using Web-based BLAST for nucleotide sequences
The Basic Local Alignment Search Tool (BLAST) available on the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>) is the key software to identify similar sequences and domain architecture of any given nucleotide/protein sequence [1].
2. Hardware: Computer system with internet connection.
3. Software: Web-browser (Internet Explorer, Mozilla Firefox, Safari, or Opera).
4. Files: In the present protocol we will continue with our previous example of ferric uptake regulator (*fur*) gene (Accession number: AF052295) from *Bradyrhizobium japonicum*.

15.2.4. Nucleotide and Protein Sequence Analysis Using SDSC Biology WorkBench

The Biology WorkBench is a Web-based utility for biologist to search and analyze nucleotide and protein sequence databases [2]. The database search is integrated with a wide variety of analysis and modeling tools. The Biology WorkBench was originally developed by the Computational Biology Group at the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, and the ongoing development of version 3.2 is occurring at the San Diego Supercomputer Center, at the University of California, San Diego. To use Biology WorkBench, users need to register at <http://workbench.sdsc.edu>.

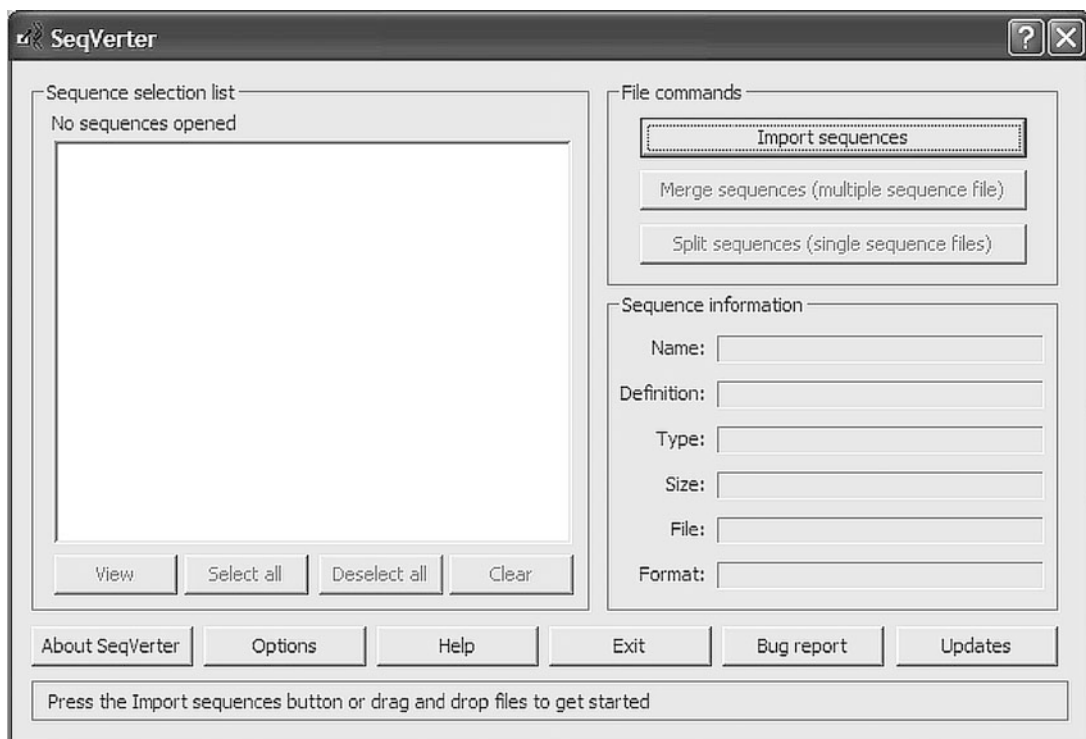


Fig. 15.7 SeqVerter interface. Sequence can be uploaded either by drag and drop utility or by selecting import sequences.

1. Hardware: Computer system with internet connection.
2. Software: Web-browser (Internet Explorer, Mozilla Firefox, Safari, or Opera).
3. Files: In the present protocol we will continue with our previous examples of ferric uptake regulator (*fur*) genes and protein sequences from various microorganisms.

15.3 Methods

15.3.1. Converting Sequence Format Using SeqVerter

1. Open the SeqVerter utility installed in the system (Fig. 15.7).
2. Load the *fur* gene in to the software by either drag and drop on the software interface or selecting import button.
3. Select the check box in front of *fur* gene as mentioned in Fig. 15.8. As the file contains only one sequence click on *Split sequences (Single sequence files)* button.
4. Select the desired location and the file format from the drop down menu of *Export single sequence file* widget (Fig. 15.9).

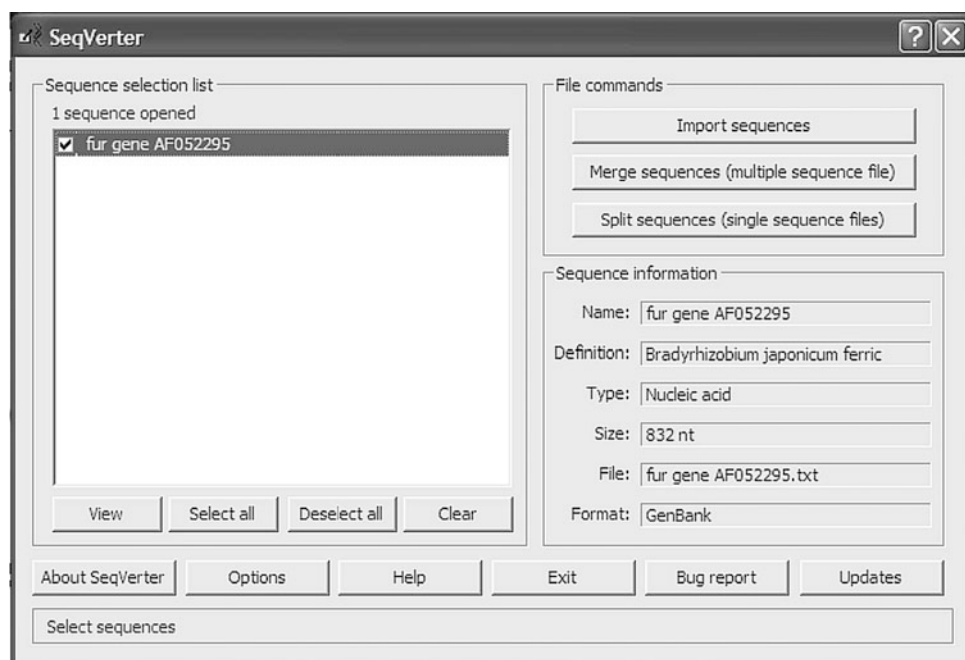


Fig. 15.8 *fur* gene AF052295 is uploaded in the software. Note the status of checkbox in front of sequence file name.

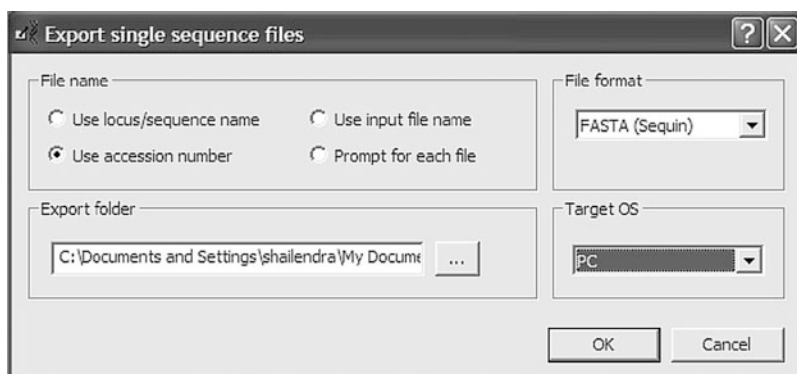


Fig. 15.9 Export single sequence files widget. Note the *Export folder* section and *File format* drop down list.

Click on OK button to convert and export the sequence format at user defined location.

5. The SeqVerter utility can also be used for multiple sequences. Figure 15.10 shows ferric uptake regulator protein (Fur protein) from ten different microorganisms.
6. Various multiple alignment formats are available in the software. This format can be selected from file format drop down menu as shown in Fig. 15.11.

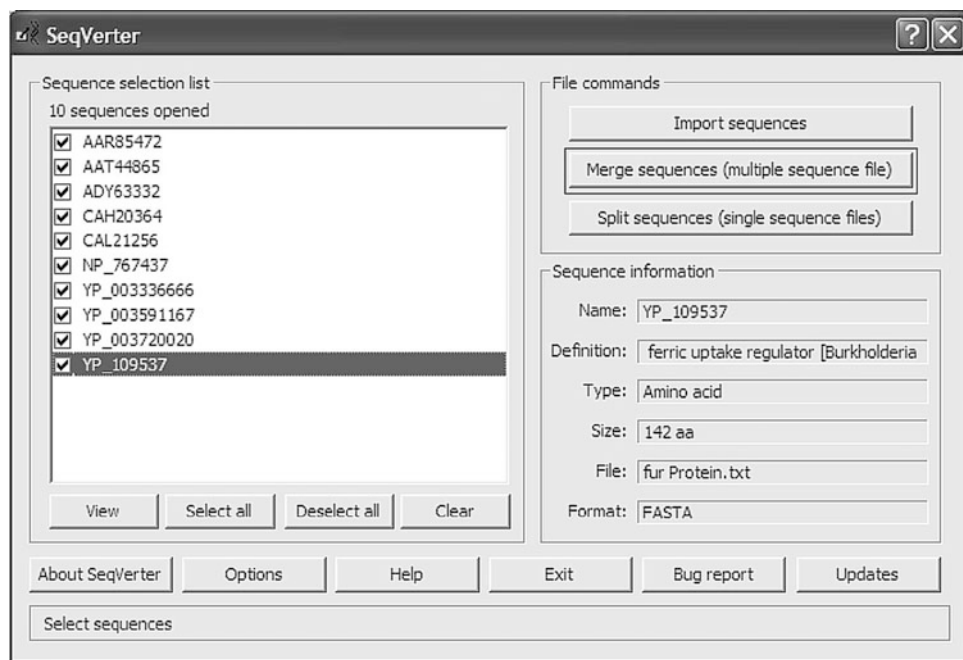


Fig. 15.10 Ferric uptake regulator proteins from ten different microorganisms are uploaded in the SeqVerter software. Note that all check boxes in front of Accession number are selected. Also note the *Merge sequences (Multiple sequence file)* button.

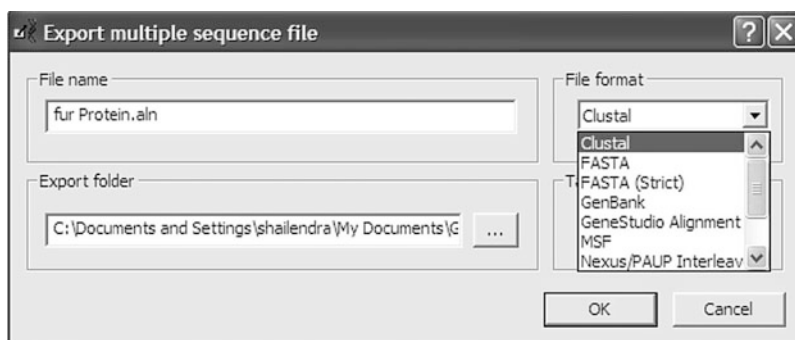


Fig. 15.11 Various multiple alignment format in the *Export multiple sequence file* widget of the software.

15.3.2. Raw Sequence Editing Tool

Biological sequences are delivered as standard chromatogram data and text files. Several software utilities are present to read and manipulate these sequence files for further bioinformatics data processing. Some of the commonly used tools are listed in Table 15.4.

One of the raw sequence editing tools FinchTV (<http://www.geospiza.com/finchtv.html>) is shown in Fig. 15.12. The tool can read DNA sequence chromatogram files. Trace peak can be scaled vertically and horizontally using *Vertical and Horizontal scale* sliders present on the left and bottom, respectively. Bases can

Table 15.4
Commonly used raw sequence editing tools

Tool	Description	Web address
FinchTV	Geospiza's FinchTV is the popular way to view DNA sequence traces. It can display entire trace in a scalable multi-pane view. It provides the utility to perform BLAST search, reverse complement sequence, and trace	http://www.geospiza.com/Products/finchtv.shtml
SeqPup	Biological sequence editor and analysis program. It includes links to network services and various external analysis programs such as clustal, cap, fastdnaml, and tacg	http://iubio.bio.indiana.edu/soft/molbio/seqpup
Sequin	A standalone utility for submitting and updating sequence entries to GenBank, EMBL, or DDBJ sequence databases	http://www.ncbi.nlm.nih.gov/Sequin/index.html
BioEdit	BioEdit is a sequence alignment editor with several sequence manipulation and analysis options and link to external analysis program	http://www.mbio.ncsu.edu/BioEdit/bioedit.html
JaMBW	Java-based Molecular Biologist WorkBench program provide utility for sequence format conversion, sequence manipulation, and various sequence analysis options, such as composition, feature viewer, ORF, isoelectric point, antigenic index, and oligo calculator	http://www.bioinformatics.org/JaMBW/
Sequencer 4.10.1	Industry standard sequence analysis software for all automated sequences including next generation sequencers. Some of the applications include de novo gene sequencing, mutation detection, systematic, etc.	http://www.genecodes.com/
CodonCode Aligner	Program for sequence assembly, contig editing, feature detection and mutation detection. The software is available for window as well as mac system	http://www.codoncode.com/aligner/index.htm
4Peaks	Sequence trace file viewing and editing system for mac OS. Software can read multiple file format, automatically translate sequence, and provide interface to add plugins to enhance functionality	http://www.mekentosj.com/science/4peaks/
PriorsEditors 1.0.10	A general editor for regulatory region analysis and transcription factor binding site discovery. De novo motif discovery program PRIORITY is bundled with the software	http://tare.medisin.ntnu.no/priorseditor/index.php

be inserted and deleted from the desired position. Sequence feature can be identified by performing online BLAST search available with the software.

15.3.3. Sequence Analysis Tools for DNA and Proteins

1. Open the Web browser with the following address <http://blast.ncbi.nlm.nih.gov>. There are two major sections for BLAST tool: (a) Basic BLAST and (b) Specialized BLAST as

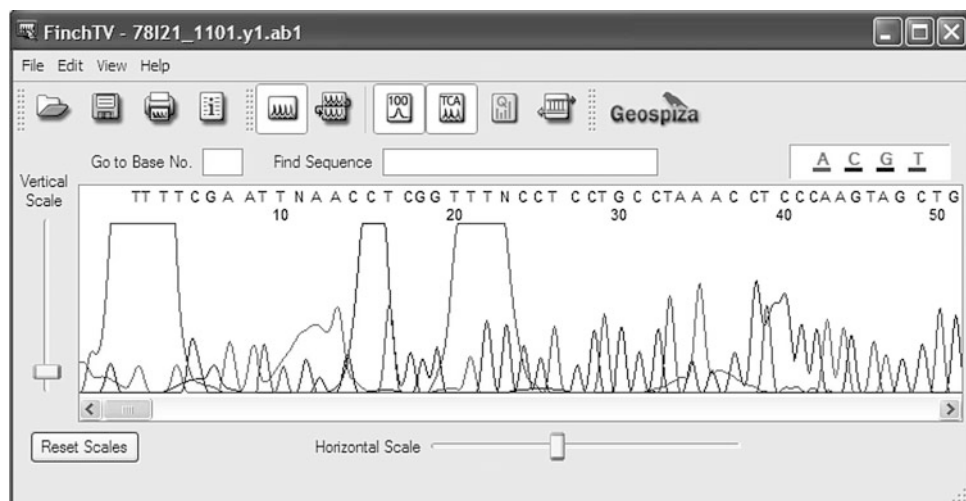


Fig. 15.12 FinchTV DNA sequence trace file reader. A sample trace file is uploaded in the software. Trace file can be edited by selecting part of the sequence and various options available in *Edit* menu.

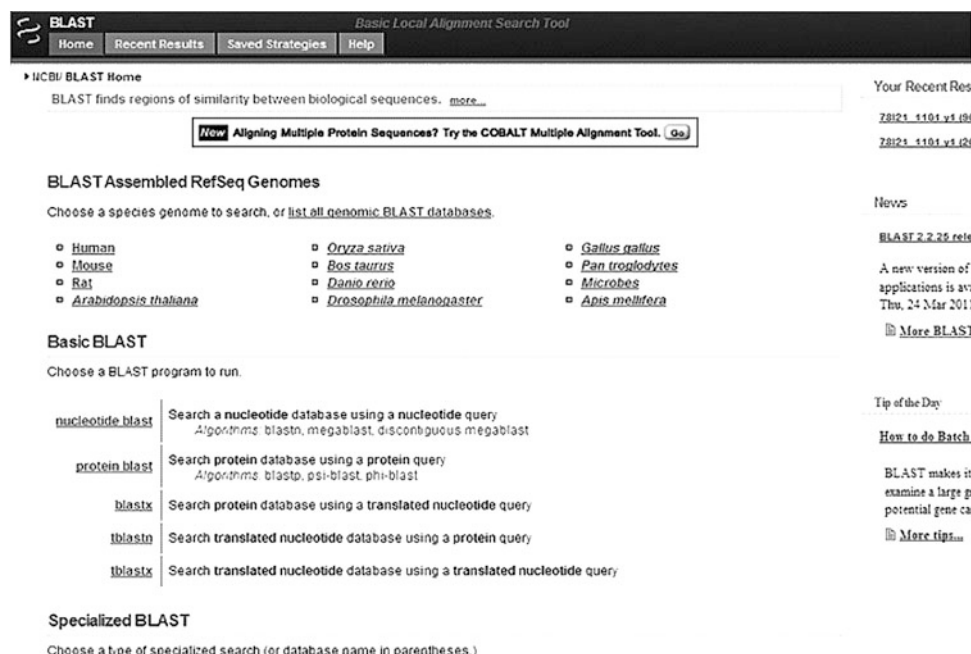


Fig. 15.13 Home page of NCBI BLAST server (<http://blast.ncbi.nlm.nih.gov>). Note five Basic BLAST programme available for nucleotide and protein sequences.

seen in Fig. 15.13. Select nucleotide BLAST under Basic BLAST tool section.

2. Paste the accession number AF052295 in the input box provided (*FASTA sequence or gi number can also be submitted for BLAST search*). BLAST tool will automatically assign a job title from the description available for the accession number.

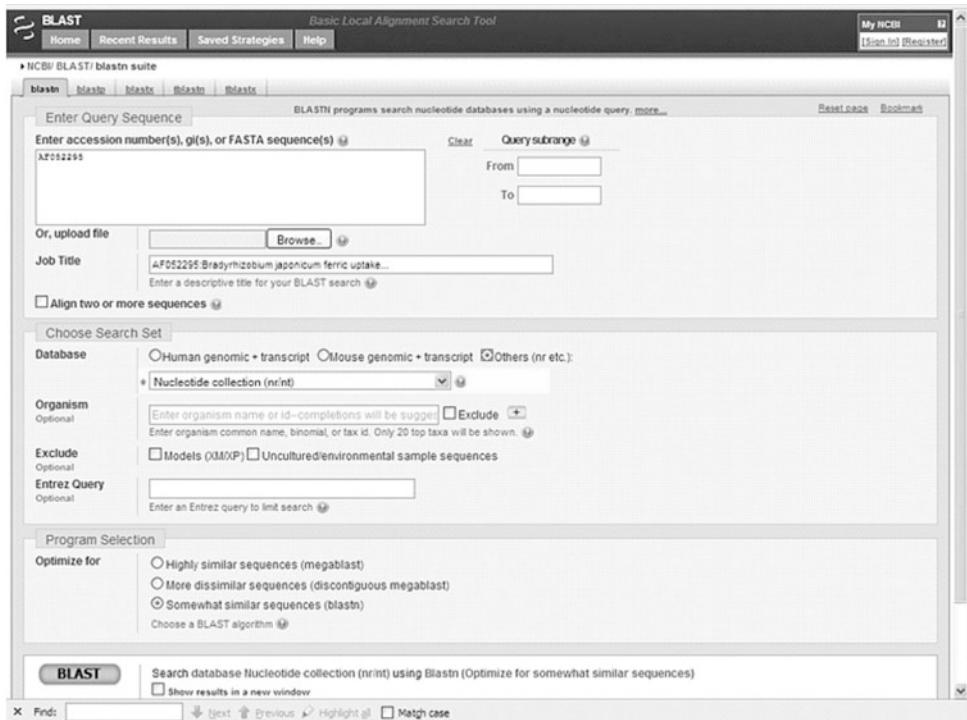


Fig. 15.14 Web interface for nucleotide BLAST search tool. Accession number for *fur* gene from *Bradyrhizobium japonicum* is provided as input. Nucleotide collection (nr/nt) database is selected for searching.

User can also provide the job title to distinguish various jobs running by the user (Fig. 15.14).

3. Query subrange can be defined if the user wants to perform BLAST search for part of the sequence only.
4. Databases are divided into three broad sections. Select “Other (nr etc.)” radio button and then select Nucleotide collection (nr/nt) from the drop down menu. List of databases is provided in Table 15.5.
5. With all the default options selected, click on the BLAST button. To increase the performance, sensitivity, and selectivity of the BLAST search, various options are available in advanced BLAST search.

15.3.3.1. BLAST Result

The basic nucleotide BLAST (blastn) output is shown in Fig. 15.15. A total of 24 BLAST hits returned by the server for the query sequence searched. Graphical summary of BLAST result is provided based on the alignment score to quickly identify the best matched sequences and their query coverage. Taxonomy and distance tree of results can be designed to find the similar sequence in diverse organisms.

Table 15.5
List of databases available for online nucleotide BLAST search

Human genomic plus transcript (Human G + T)
Mouse genomic plus transcript (Mouse G + T)
Nucleotide collection (nr/nt)
Reference mRNA sequences (refseq_rna)
Reference genomic sequences (refseq_genomic)
NCBI Genomes (chromosome)
Expressed sequence tags (est)
Nonhuman, non-mouse ESTs (est_others)
Genomic survey sequences (gss)
High-throughput genomic sequences (HTGS)
Patent sequences (pat)
Protein Data Bank (pdb)
Human ALU repeat elements (alu_repeats)
Sequence tagged sites (dbsts)
Whole-genome shotgun reads (wgs)
Environmental samples (env_nt)

15.3.4. Nucleotide and Protein Sequence Analysis Using SDSC Biology WorkBench

1. Point the Web browser to <http://workbench.sdsc.edu>. Click on the links to enter into Biology WorkBench, if already registered on the server or else click on *register* link as shown in Fig. 15.16.
2. Basic structure of Biology WorkBench can be defined into three major sections. These include (1) Tool Sets, (2) Sequence, and (3) Tools as shown in (Fig. 15.17).

Tool sets currently include:

- (a) Session Tools
- (b) Protein Tools
- (c) Nucleic Tools
- (d) Alignment Tools
- (e) Structure Tools (Alpha)
- (f) Report Bugs

The WorkBench views sequences as objects on which it can perform tasks. These sequences can be imported either through your machines or can be searched with different biological databases. The Protein Tools take protein sequences as inputs. The Nucleic Tools take nucleotide

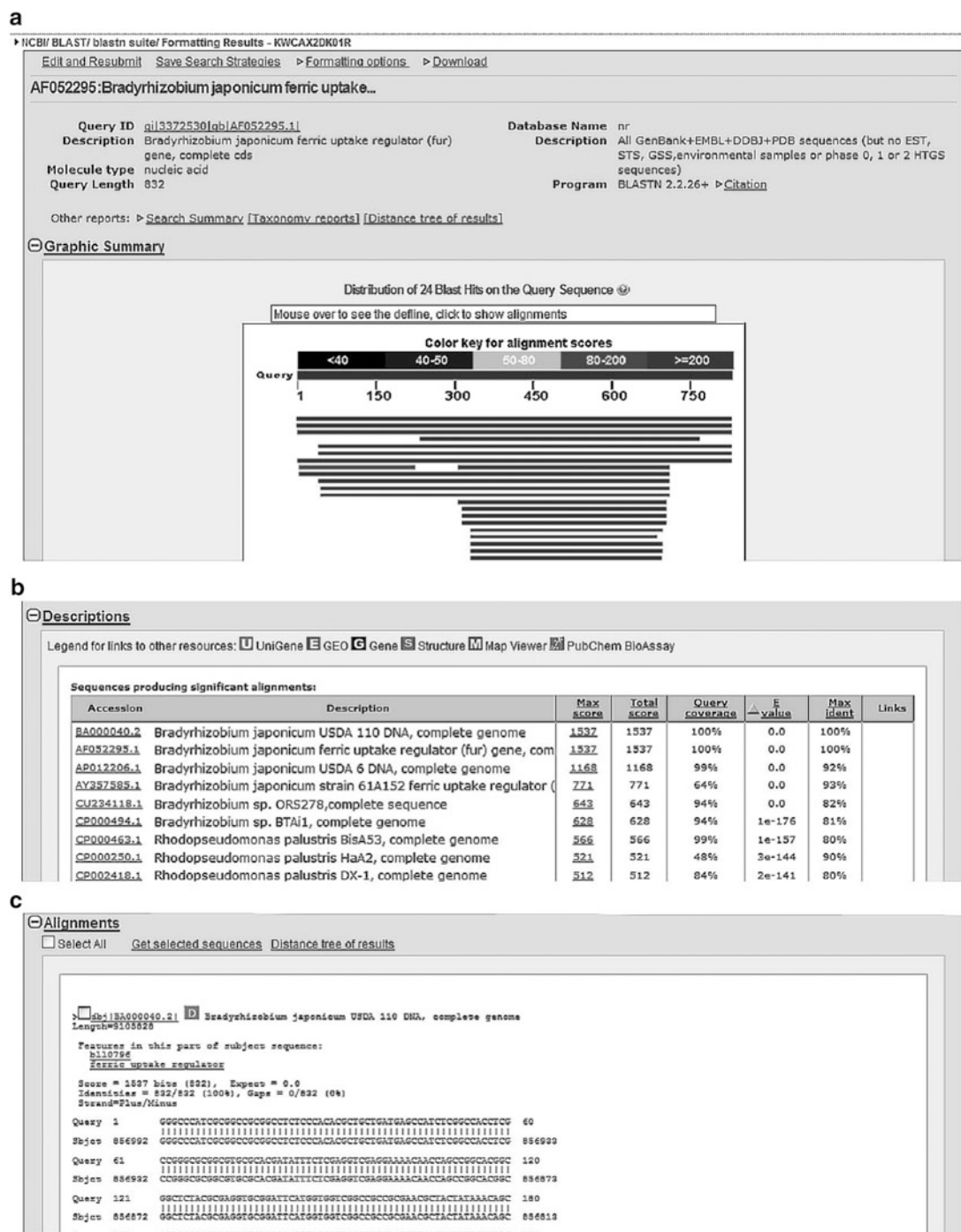


Fig. 15.15 Blastn output. (a) Distribution of hits and color-coded graphical result. (b) List of related sequences, accession number, description, score, E-value, and database linkages. (c) The detailed pairwise alignment between the query sequence and the database sequence.

SDSC
SAN DIEGO SUPERCOMPUTER CENTER
Biology WorkBench

The **Biology WorkBench** is a web-based tool for biologists. The WorkBench allows biologists to search many popular protein and nucleic acid sequence databases. Database searching is integrated with access to a wide variety of analysis and modeling tools, all within a point and click interface that eliminates file format compatibility problems.

First time users: please [register](#) for a free account.

Click to Enter the Biology Workbench 3.2

Forgotten Password: there are two ways this can be fixed, once we verify you own the account in question. One is for you to register for a new account, and we can transfer the data from your old account. The second is for us to reset your old account, and then you can register for it again - your old data will still be available. Please mail bwbhelp@sdsc.edu and let us know which option you prefer.

Announcing the Next Generation Biology Workbench

The Next Generation Biology Workbench (NGBW) is now available for public use, in its v.1.5 full production release. Please visit the NGBW at www.ngbw.org, and provide feedback. Some of the new features that the NGBW contains are:

- Imports data from the original Biology Workbench
- Full support for modeling and visualization of biological structures, including an integrated tool (Sirius)

Fig. 15.16 Home page of SDSC Biology WorkBench at <http://workbench.sdsc.edu>.

sequences as inputs and the Alignment Tools take aligned sequences as inputs. Alignment between nucleotide/protein sequences can be performed using tools in either the Protein or Nucleic Tools set. The output of the programs will be automatically submitted into the Alignment tools, where you can perform operations on aligned sequences.

3. Click on the *Session Tools* button to start the default session of the Biology WorkBench. You can rename and save the session so that you can work on the same sets of sequences in future.
4. Click on the *Nucleic Tools* button. The programs associated with the nucleotide sequences analysis will be listed in the dropdown menu as shown in the (Fig. 15.18). Select the *Add New Nucleotide Sequence* and click on *Run* button.
5. Sequence can be uploaded either by selecting *Browse* button and then *Upload File* button (Fig. 15.19). Sequence can also be entered manually on the server by providing *Label* and *Sequence data*.
6. Click on the *Save* button to save the sequences on the current session for further analysis. Various sequence analysis tools are available on WorkBench. Analysis can be performed by selecting the checkbox in front of each sequence and selecting the appropriate tool from the list followed by selecting *Run* button (Fig. 15.20). List of various tools available for nucleotide sequences is provided in Table 15.6.

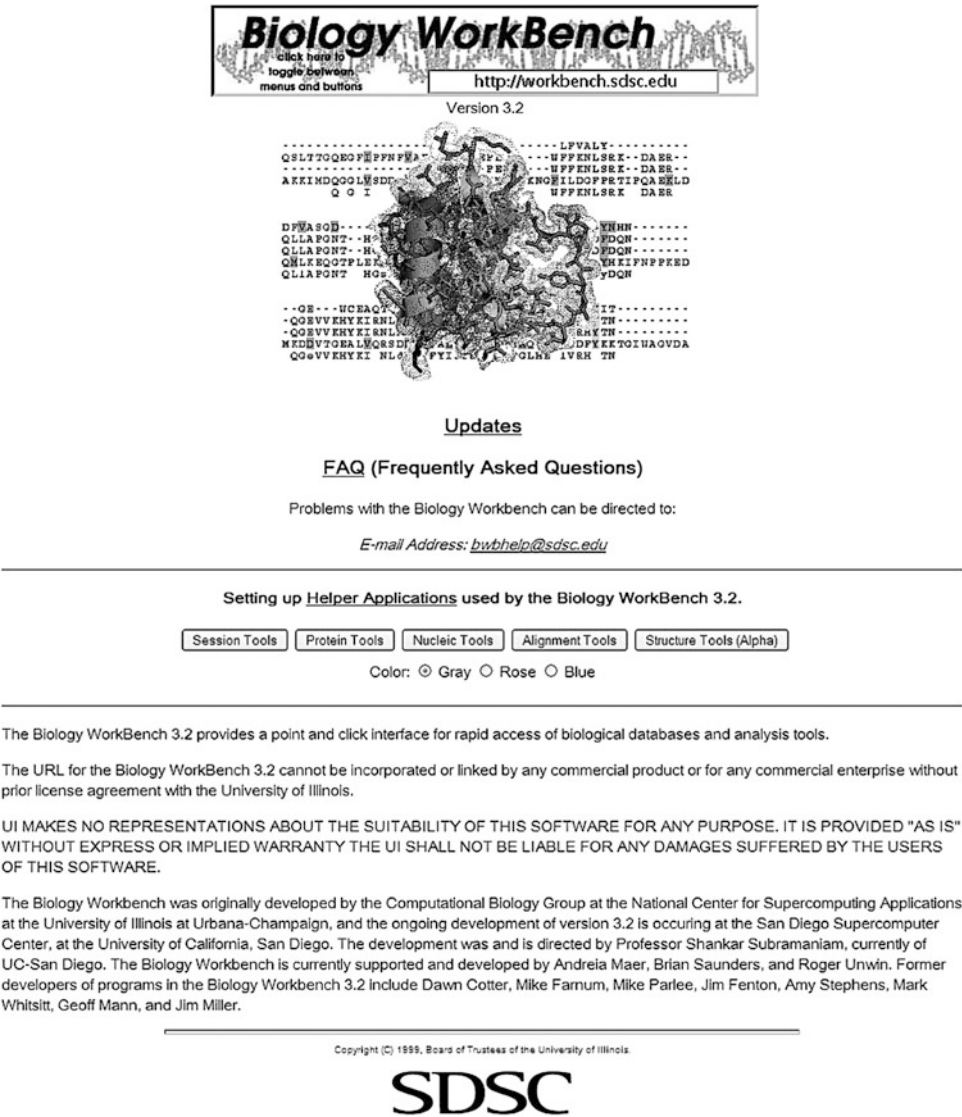


Fig. 15.17 SDSC Biology WorkBench. Note five buttons to initiate the work.

15.3.4.1. Calculation of Nucleotide Sequence Statistics

First, we will check the composition of *Bradyrhizobium japonicum fur* gene, complete cds using NASTATS tool. Select the check box in front of sequence and select *NASTATS Tool* from the list under *Nucleic Tools Set* and click on *RUN* button. The output of the NASTATS tool is shown in Fig. 15.21.

15.3.4.2. Align Two Sequences Using BL2SEQ Utility

In the present protocol, we will compare *fur* gene sequences from two different organisms using BLAST utility available on WorkBench. Select the check box in front of *fur* gene from *Bradyrhizobium japonicum*. Select the BL2SEQ utility under Nucleic Tools set. Click on *Run* button to submit the job. In the next screen,

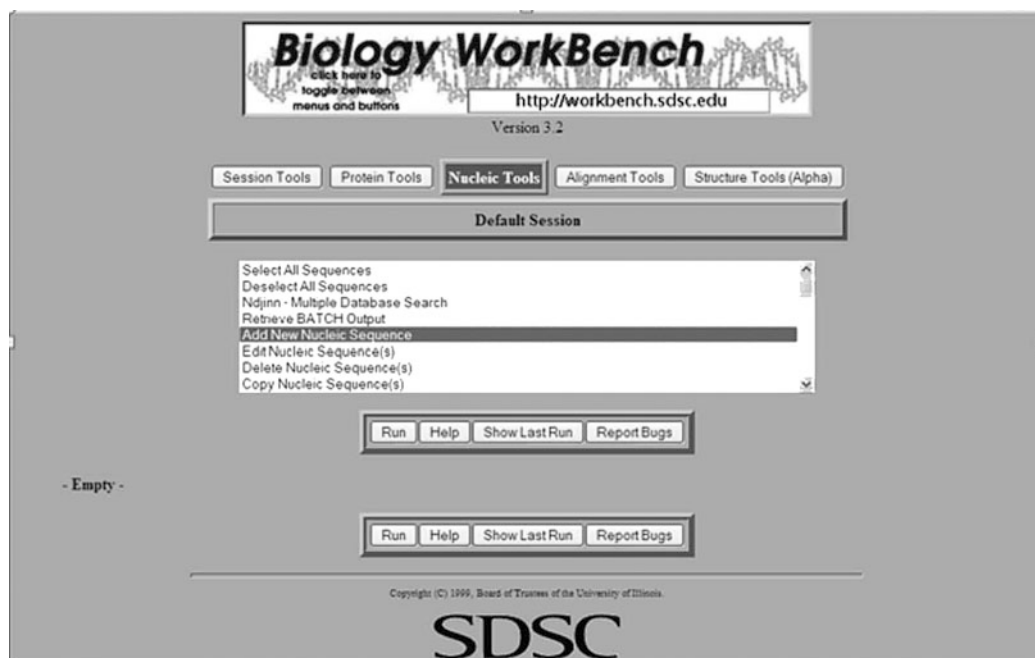


Fig. 15.18 Sequence can be uploaded to the WorkBench by selecting *Add New Nucleotide Sequence* tool. Note the Empty tag, which shows that currently no sequence is associated with the WorkBench.

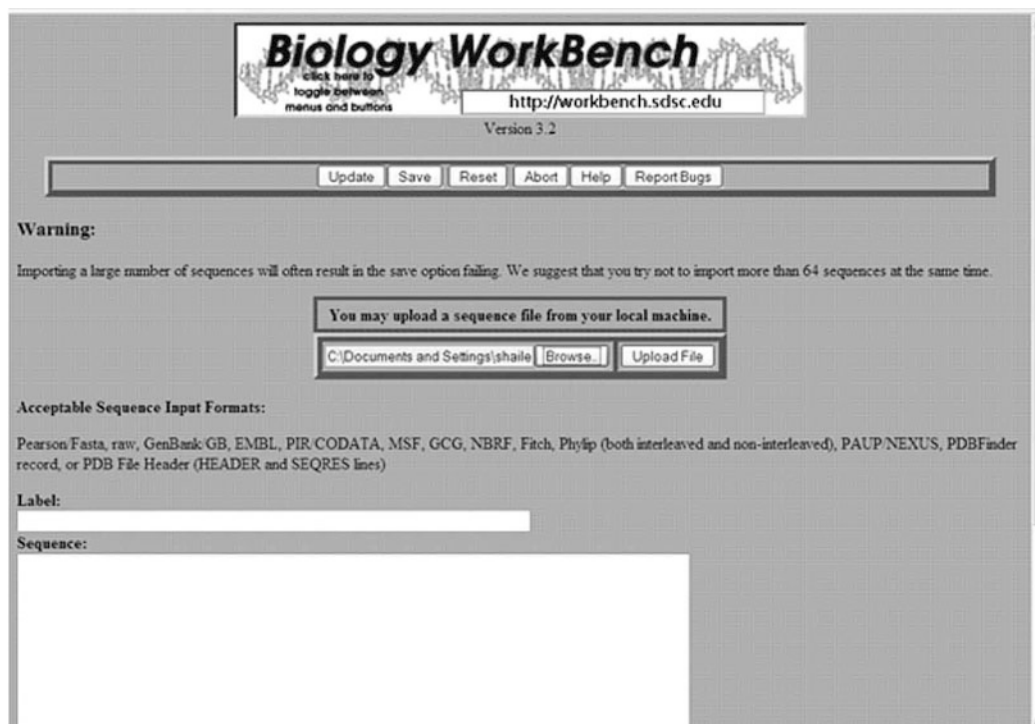


Fig. 15.19 Snapshot of Upload sequence utility on Biology WorkBench. Note the acceptable sequence input formats available with the WorkBench.

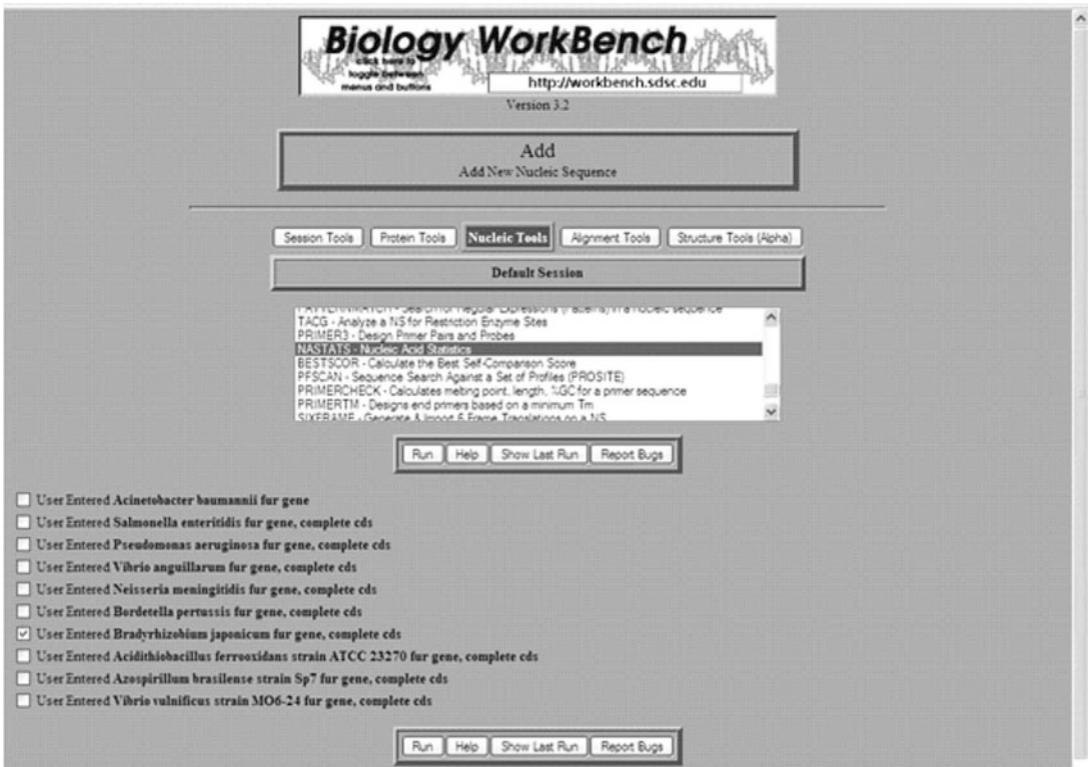


Fig. 15.20 Total ten nucleotide sequences from different organisms for *Fur* gene is uploaded on the current session. Note the check box in front of each sequence and the various tools associated with Nucleic Tools set.

select from the list of organism *Azospirillum brasilense* [you can also select more than one sequence for comparison]. Enter the *Expectation value* 10 and click on *Submit* button. The output generated by the tool is shown in Fig. 15.22. A small local alignment region between the sequences from two organisms is shown as pairwise alignment. The position of aligned fragment, identity, and similarity score is also returned.

15.3.5. Nucleotide Sequence Similarity Searching Using BLASTN Tool on Biology WorkBench

Select the check box in front of *fur* gene from *Bradyrhizobium japonicum*. Select *BLASTN* tool from the Nucleic Tools set and click on *Run* button. *BLASTN* compares a single DNA sequence (query sequence) to all sequences in a database (library sequences) to find those library sequences having the greatest similarity to the query sequence. *BLASTN* uses a heuristic algorithm to allow faster database searching. On the following screen, select databases to be searched (A maximum of 16 databases can be searched in single run). Search can be performed with both the forward and backward strand of the query sequence (Fig. 15.23). Select GenBank Bacterial Sequences [part 1] database from the list and click on submit button. The output generated by the Biology WorkBench is shown in Fig. 15.24. We can also fetch the details of the

Table 15.6
List of various utilities available with Nucleic Tool Set of
Biology WorkBench

Select All Sequences
Deselect All Sequences
Ndjinn—Multiple Database Search
Retrieve BATCH Output
Add New Nucleic Sequence
Edit Nucleic Sequence(s)
Delete Nucleic Sequence(s)
Copy Nucleic Sequence(s)
View Nucleic Sequence(s)
Download Nucleic Sequence(s)
View Database Records of Imported Sequences
BL2SEQ—Compare nucleotides to each other with BLAST
BL2SEQX—Compare a nucleotide to protein sequences with BLAST
BLASTN—Compare a NS to a NS DB
BLASTX—Compare a PS-Derived-from-NS to a PS DB
TBLASTX—Compare a translated NS to a translated DB
FASTA—Nucleic Acid Sequence Comparisons (NS or DB)
FASTX—Compare Translated NS to PS DB
FASTY—Compare Translated NS to PS DB
SSEARCH—Smith-Waterman Local Alignment
CLUSTALW—Multiple Sequence Alignment
CLUSTALWPROF—Align Sequences to Existing Alignment (Profile)
ALIGN—Optimal Global Sequence Alignment
LALIGN—Calculate Optimal Local Sequence Alignments
LFASTA—Calculate Local Sequence Alignments (Heuristic)
PATTERNMATCHDB—Search for Regular Expressions (Patterns) in a nucleic sequence DB
PATTERNMATCH—Search for Regular Expressions (Patterns) in a nucleic sequence
TACG—Analyze a NS for Restriction Enzyme Sites
PRIMER3—Design Primer Pairs and Probes
NASTATS—Nucleic Acid Statistics
BESTSCOR—Calculate the Best Self-Comparison Score

(continued)

Table 15.6
(continued)

PFSCAN—Sequence Search Against a Set of Profiles (PROSITE)
PRIMERCHECK—Calculates melting point, length, %GC for a primer sequence
PRIMERTM—Designs end primers based on a minimum Tm
SIXFRAME—Generate and Import 6 Frame Translations on a NS
REVCOMP—Generate Reverse Complement of NS
RANDSEQ—Randomize a Sequence

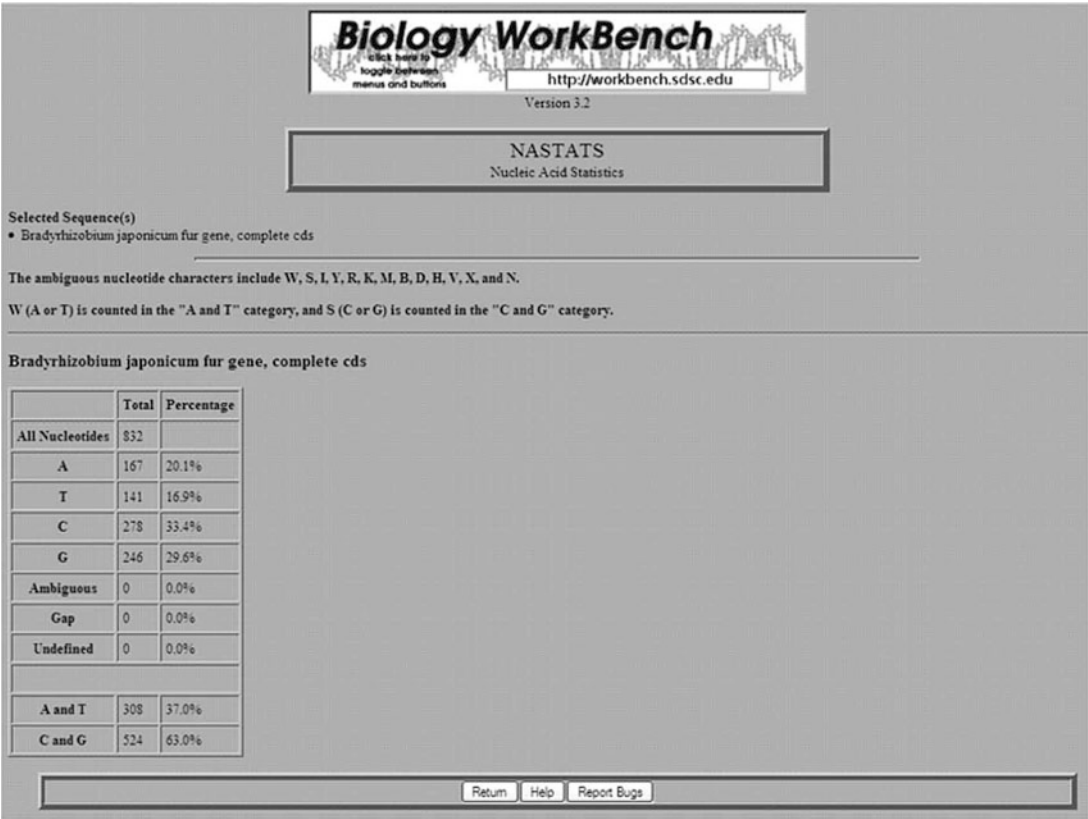


Fig. 15.21 Result screen of NASTATS Tool for *Bradyrhizobium japonicum fur* gene. Note the nucleotide percentage and GC content of the gene.

database sequence by selecting check box in front of each 1-line description followed by Show Records button. The output result is arranged in the decreasing order of bit score, which is defined as:

$$S'(bits) = [\lambda \times S(raw) - \ln K] / \ln 2$$

where lambda and K are Karlin–Altschul parameters.

```

Bradyrhizobium japonicum fur gene, complete cds (User Entered)
- and -
Azospirillum brasilense strain Sp7 fur gene, complete cds (User Entered)

Query:      >Bradyrhizobium japonicum fur gene, complete cds
            Length = 832
Reference: Query= Bradyrhizobium japonicum fur gene, complete cds
            (832 letters)

>Azospirillum brasilense strain Sp7 fur gene, complete cds
            Length = 638

Score = 26.3 bits (13), Expect = 0.006
Identities = 16/17 (94%)
Strand = Plus / Plus

Query: 395 catcccgagctggagga 411
      |||
Sbjct: 211 catcccgagctggagga 227

Score = 22.3 bits (11), Expect = 0.10
Identities = 11/11 (100%)
Strand = Plus / Plus

Query: 591 aggtgatcgag 601
      |||
Sbjct: 211 aggtgatcgag 221

```

Fig. 15.22 BL2SEQ output for *Bradyrhizobium japonicum* and *Azospirillum brasilense* fur gene.

Biology WorkBench
http://workbench.sdsc.edu
Version 3.2

BLASTN
Compare a NS to a NS DB

Selected Sequence(s)
• Bradyrhizobium japonicum fur gene, complete cds

Enter the required information below; then press "Submit" to run the job.

Query strands to search: Both

Choose from 1 up to 16 Databases:
☐ DICTYOSTELIUM ORFs used for annotations
☐ DICTYOSTELIUM all HMM predicted ORFs
☒ GenBank (nr)
☐ GenBank Bacterial Sequences (part 1)
☐ GenBank Bacterial Sequences (part 2)
☐ GenBank Environmental Survey Sequences (part 1)
☐ GenBank Environmental Survey Sequences (part 2)
☐ GenBank High Throughput cDNA Sequencing Series
☐ GenBank Invertebrate Sequences (part 1)
☐ GenBank Invertebrate Sequences (part 2)

Expectation value: 10
use decimal or scientific (for example 1.0e-5) notation

Disable DUST filtering of query string: ☐

1-line descriptions: 100

Alignments: 50

Submit Show Last Run Abort Set Default Reset Help Report Bugs

☐ Run as batch

Citation
Algorithm Citation:
Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1990), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 18:3500-3509.

Fig. 15.23 BLASTN input screen on WorkBench. Expectation value, 1-line description, and no. of alignments returned by the server can be adjusted.

Output order: **Aligned** (the order of sequences in the alignment output)
 Guide tree display: **Rooted and Unrooted Trees** (rooted tree is inferred from unrooted tree)

PAIRWISE ALIGNMENT PARAMETERS

Alignment method: **Accurate**

Accurate method parameters

Weight matrix: **ClustalW(1.6)**
 Gap open penalty: **15.00** (0.0 - 100.0)
 Gap extension penalty: **6.66** (0.0 - 10.0)

Fast method parameters

K-tuple Size: **2** (1 - 500)
 Gap penalty: **5** (1 - 500)
 Top diagonals: **4** (1 - 50)
 Window size: **4** (1 - 50)

MULTIPLE ALIGNMENT PARAMETERS

DNA transitions weight: **0.5** (0.0 - 1.0)
 Weight matrix: **ClustalW(1.6)**
 Use negative matrix: **No**
 Gap open penalty: **15.00** (0.0 - 100.0)
 Gap extension penalty: **6.66** (0.0 - 10.0)
 Delay divergent sequences: **30** (0 - 100)

☐ **Run as batch**

Citation

Algorithm Citation:
 Higgins D.G., Bleasby, A.J. and Fuchs, R. (1992) CLUSTAL V: improved software for multiple sequence alignment. *Computer Applications in the Biosciences (CABIOS)*, 8(2):189-191.
 Thompson J.D., Higgins D.G., Gibson T.J. "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Res.* 23:4031-4044 (1994).

Fig. 15.25 Snapshot of the screen to adjust various parameters in *CLUSTALW—Multiple Sequence Alignment Tool*.

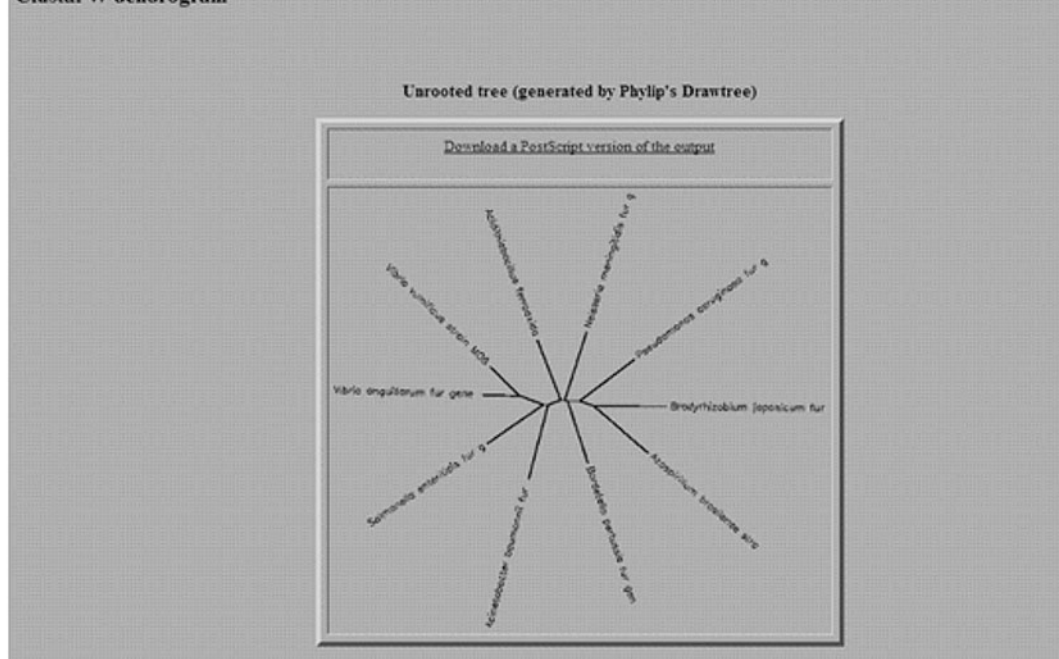
Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

15.3.6. Multiple Sequence Alignment Using CLUSTALW Utility on Biology WorkBench

The Clustal programs are widely used for carrying out automatic multiple alignments of nucleotide or amino acid sequences. The most familiar version is ClustalW [3]. Select the checkbox in front of all the ten *fur* genes from various microorganisms uploaded on the Biology WorkBench server in the earlier protocols. Select the *CLUSTALW—Multiple Sequence Alignment* option from the Nucleic Tools set. Click on the *Run* button. On the next screen, various pairwise and multiple alignment parameters can be adjusted. Select *Aligned* from *Output order* dropdown list and *Rooted and Unrooted Trees* from *Guide tree display* dropdown list. From the *Multiple Alignment Parameters* section, select the *ClustalW(1.6)* option out of *Weight Matrix* dropdown list (Fig. 15.25). Click on the *Submit* button to run the *CLUSTALW* tool. The output is shown in Fig. 15.26a–c. Other multiple alignment tools available in the public domains are listed in Table 15.7 and phylogenetic analysis tools are listed in Table 15.8. The alignment file generated by the *CLUSTALW* program can be imported on the Biology WorkBench by selecting *Import Alignment(s)* button. The ClustalW aligned file can be further processed by various tools in the *Alignment Tools set*.



b Clustal W dendrogram



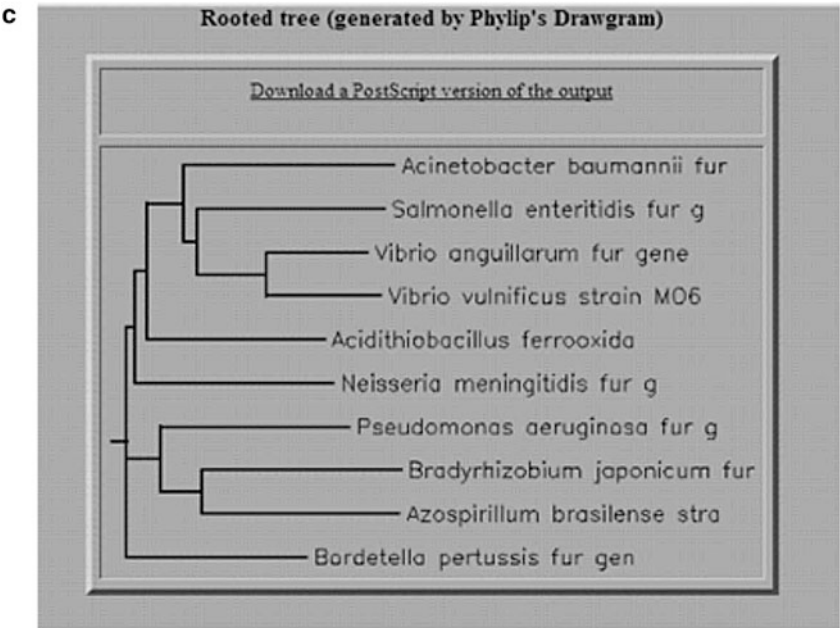


Fig. 15.26 (a) The multiple alignment screen for ten *fur* genes from different organisms. The aligned regions are highlighted in *blue color*. (b) Unrooted tree generated by Phylip's Drawtree program between multiple aligned sequences. (c) Rooted tree generated by Phylip's Drawtree program. Note the similar organisms based on the distances of *fur* genes.

Table 15.7
List of frequently used multiple sequence alignment tools

Name	Description	Web address
MAFFT	Recent developments in the MAFFT multiple sequence alignment program	http://align.bmr.kyushu-u.ac.jp/mafft/software/
Clustal-W	ClustalW2 is a general purpose global multiple sequence alignment program for DNA or proteins	http://www.ebi.ac.uk/clustalw
PCMA	Fast and accurate multiple sequence alignment based on profile consistency	ftp://iole.swmed.edu/pub/PCMA/
ProAlign	A hidden Markov model for progressive multiple alignment	http://ueg.ulb.ac.be/ProAlign/
MAVID	Constrained ancestral alignment of multiple sequences	http://baboon.math.berkeley.edu/mavid
ABA	A novel method for multiple alignment of sequences with repeated and shuffled elements	http://nbc.sdsu.edu/euler
POA	Combining partial order alignment and progressive sequence alignment increases alignment speed and scalability to very large alignment problems	http://www.bioinformatics.ucla.edu/poa/
DIALIGN	DIALIGN-T: an improved algorithm for segment-based multiple sequence alignments	http://dialign.gobics.de/chaos-dialign-submission
NRAlign	Improving accuracy of multiple sequence alignment algorithms based on alignment of neighboring residues	http://faculty.cs.tamu.edu/shsze/nralign

Table 15.8
List of important phylogenetic analysis tools

Name	Description	Web address
Phylip	Is a free package of programs for inferring phylogenies. It is distributed as source code, documentation files, and a number of different types of executables	http://evolution.genetics.washington.edu/phylip.html
Phylogeny.fr	Is a simple to use Web service dedicated to reconstructing and analyzing phylogenetic relationships between molecular sequences. Phylogeny (PhyML, MrBayes, TNT, BioNJ), tree viewer (Drawgram, Drawtree, ATV)	http://www.phylogeny.fr/version2_cgi/index.cgi
PHYML	Is a simple, fast, and accurate algorithm to estimate maximum likelihood phylogenies from DNA and protein sequences	http://atgc.lirmm.fr/phyml/
ProtTest	Estimates the empirical model of amino acid substitution that fits the data best among 64 candidate models	http://darwin.uvigo.es/
POWER	To carry out phylogenetic analysis on most programs of PHYLIP package repeatedly	http://power.nhri.org.tw/power/home.htm
CVTree	Constructs whole-genome-based phylogenetic trees without sequence alignment by using a Composition Vector (CV) approach	http://tlife.fudan.edu.cn/cvtree/

Selected Sequence(s)
• Bradyrhizobium japonicum for pma, complete cds

Submit Abort Reset Help Report Bugs

Run as batch

You may change the name of this sequence

Sequence ID: Bradyrhizobium japonicum for pma, complete cds

Database ID: User Entered

1 BRDITGATG GCGCGGAC CTTCTGCA CTTCTGAT GACGCTCT BRDCTGTT
41 CTGAGGCG GTCGCGCA ATATTCTT AATTGAGA AAGACATG CTGACGAC
121 AGCTTACG GAGTGGCA TGTGATTA TGGGCTGT GACAGCTG TGAAGAC
181 CAGACGGA AATTGACG GCTCTCTA TGGGCTTA CTGTGATA CATTGATG
241 AAGAGCTG CTGAGGCA GACGCTCT GACGCTCA AACTCTTT CTGCTGTA
301 AGCTTACG GAGTGGCA GCTGCTCT CAGGCTCA GCTGCTCA AAGGCTGT
361 CTTGCTCT GCTGCTCT CAGAGGAT GAGCTGCT GACTTGAG AATTGATG
421 GCTGCTCT GCTGCTCT AAGAGCTT GCTGCTCT GCTGCTCA GCTGCTCT
481 GCTGCTCT GCTGCTCT TGAAGCTA TACTTCTG GAGGAGCT GCTGCTCA
541 GACGCTCT GACGCTCT AAGAGCTT CATGATCT GCTGCTCA AATTGATG
601 GCTGCTCT AAGAGCTT AAGGCTCT GCTGCTCT GCTGCTCA TGTGATTA
661 GCTGCTCT GCTGCTCT AACTTCTT GCTGCTCT GCTGCTCA AACTGATG
721 CAGGCTCT GCTGCTCT TGTGATTA CTTGCTCT GCTGCTCT AACTGATG
781 GACTGCTCT GCTGCTCT GAGGCTCT GACTGCTCT GCTGCTCA TC

Sequence Region Selection

Included Region	Start: 1 End: 832	Enter start and end residues of range to select primers from. (Default is entire sequence)
Target Regions	Start: End: 1: <input type="text"/> <input type="text"/> 2: <input type="text"/> <input type="text"/> 3: <input type="text"/> <input type="text"/> 4: <input type="text"/> <input type="text"/>	Enter start and end residues of ranges to be targeted by primers. Acceptable primer pairs will flank these regions. You may select up to 4 target regions. (Default is empty)
Excluded Regions	Start: End: 1: <input type="text"/> <input type="text"/> 2: <input type="text"/> <input type="text"/> 3: <input type="text"/> <input type="text"/> 4: <input type="text"/> <input type="text"/>	Enter start and end residues to be excluded from primers. Acceptable primer pairs will not overlap these regions. You may select up to 4 excluded regions. (Default is empty)
Left Primer	<input type="text"/>	Enter sequence of leftmost primer to be used (optional). This primer will be used to identify a rightmost primer meeting the criteria set below. No other left primers will be considered. (Default is no left primer)
Right Primer	<input type="text"/>	Enter sequence of rightmost primer to be used (optional). This primer will be used to identify a leftmost primer meeting the criteria set below. No other right primers will be considered. (Default is no right primer)
Number of Primers to Display	5	Number of primer pairs to be displayed. Increasing this number will increase running time. (Default is 5 primer pairs)

Fig. 15.27 Some of the Primer3 input parameters for designing optimal primer for the selected sequence on Biology WorkBench.

Table 15.9
List of commonly used primer/probe designing tools

Name	Description	Web address
AutoPrime	Primer design for real-time PCR measurement of eukaryotic gene expression	http://www.autoprime.de/AutoPrimeWeb
CODEHOP	COnsensus-DEgenerate Hybrid Oligonucleotide Primers designed from protein multiple sequence alignments	http://bioinformatics.weizmann.ac.il/blocks/codehop.html
ExonPrimer	Design intronic primers for PCR amplification of exons. Input needed: a cDNA and the corresponding genomic sequence	http://ihg.gsf.de/ihg/ExonPrimer.htm
NetPrimer	Java applet for primer design	http://www.premierbiosoft.com/netprimer/
Primer3	Utility for locating oligonucleotide primers for PCR amplification of DNA sequences	http://frodo.wi.mit.edu/primer3/
PrimerX	Automated design of mutagenic primers for site-directed mutagenesis	http://www.bioinformatics.org/primerx/
Primo Pro	PCR Primer Design	http://www.changbioscience.com/primo
Web Primer	Primer design and sets for amplifying yeast ORFs	http://www.yeastgenome.org/cgi-bin/web-primer

similar melting temperatures since annealing in a PCR occurs for both simultaneously [4]. Primers should not easily anneal with other primers in the mixture; this phenomenon can lead to the construction of “primer dimer” products contaminating the mixture. Primers should also not anneal strongly to themselves, as internal hairpins and loops could hinder the annealing with the template DNA. Desired characteristics of an automated DNA sequencing primer design are:

- Based on accurate sequence
- Melting temperature (T_m): 52–65 °C
- Absence of self-hybridization
- Absence of significant hairpin formation (>3 bp)
- Lack of secondary priming sites
- Low specific binding at the 3' end (i.e., lower GC content to avoid mispriming)

Biology WorkBench provides interface to Primer3 software, which is a widely used program for designing PCR primers. Primer3 predicts primer based on following criteria:

- (a) Oligonucleotide melting temperature, size, GC content, and primer-dimer possibilities

- (b) PCR product size
- (c) Positional constraints within the source sequence
- (d) Miscellaneous other constraints

All of these criteria are user specifiable as constraints, and some are specifiable as terms in an objective function that characterizes an optimal primer pair. Select the *fur* gene from *Bradyrhizobium japonicum*. Choose the Primer3 Tool from the *Nucleic Tools Set*. Click on *Run* button. Next screen (Fig. 15.27) will appear to provide *Sequence Region Selection* and *Primer Selection Criteria* to design the best primer for the selected sequence. The Primer3 output on Biology WorkBench is shown in Fig. 15.28. Both left and right primer sets are shown in the figure. The output include start position of the primer, their length, melting temperature, GC %, any (self-complementarity score of primer: taken as a measure of its tendency to anneal itself or from secondary structure), 3' (self-complementarity: taken as a measure of its tendency to form a primer-dimer with itself) and the primer sequence are provided in the result. The left and right primer sequence can also be uploaded on the Biology WorkBench session simply by pressing *Import Sequence(s)* button. Other commonly used primer designing tools are listed in Table 15.9.

15.3.8. Bioinformatics Tools for Metagenomics

Metagenomics is a rising field in which the control of genomic analysis (the analysis of the entire DNA in an organism) is useful to whole communities of microbes, bypassing the requirement to isolate and culture individual microbial species. By permitting the direct investigation of bacteria, viruses, and fungi irrespective of their culturability and taxonomic identities, metagenomics has altered microbiological theory and methods and has challenged the classical concept of species. This latest field of biology has proven to be rich and comprehensive and is making significant contributions in numerous areas including ecology, biodiversity, bioremediation, bioprospection of natural products, and in medicine. Various bioinformatics tools have been developed for metagenomics analysis. Some of them are listed in Table 15.10.

15.3.9. Bioinformatics Tools for Metabolomics

Metabolomics is the “organized study of the unique chemical fingerprints that exact cellular processes leave behind,” the study of their small-molecule metabolite profiles. The metabolome represents the group of all metabolites in a biological cell, tissue, organ, or organism, which are the end products of cellular processes. One of the challenges of systems biology and functional genomics is to incorporate proteomic, transcriptomic, and metabolomic information to give a more complete picture of living organisms. Some of the tools have been listed in Table 15.11.

Table 15.10
List of important metagenomics analysis tools

Name	Description	Web address
MEtaGenome Analyzer	Taxonomic analysis	http://ab.inf.uni-tuebingen.de/software/megan/
Integrated Microbial Genomes	Microbial Genome Data Management and Analysis Systems	http://img.jgi.doe.gov/
PyroTagger	Classify multiplexed amplicon pyrosequence data from any region of the SSU rRNA gene when provided with barcode and primer sequences	http://pyrotagger.jgi-psf.org/cgi-bin/index.pl
CLaMS	Sequence composition-based classifier for metagenomic sequences	http://clams.jgi-psf.org/
FAMeS	Simulated data sets to evaluate the fidelity of metagenomic processing methods	http://fames.jgi-psf.org/
GOLD	Comprehensive access to information regarding complete and ongoing genome projects, as well as metagenomes and metadata	http://www.genomesonline.org/

Table 15.11
List of important metabolomics analysis tools

Name	Description	Web address
The Human Metabolome Database	Information about small molecule metabolites found in the human body	http://www.hmdb.ca/
Fiehn Laboratory	Identification and quantification of all metabolites in a given biological situation	http://fiehnlab.ucdavis.edu/
MMCD	High-throughput NMR and MS approaches to the identification and quantification of metabolites present in biological samples	http://mmcd.nmr.fam.wisc.edu/
MeT-RO	Establish a critical mass of resources that can be applied to plant and microbial metabolomics	http://www.metabolomics.bbsrc.ac.uk/MeT-RO.htm
PRiMe	Metabolomics and Transcriptomics as systems for understanding life	http://prime.psc.riken.jp/
METLIN	A repository of metabolite information as well as tandem mass spectrometry data	http://metlin.scripps.edu/
MZmine	Mass spectrometry data processing, with the main focus on LC-MS data	http://mzmine.sourceforge.net/index.shtml
MathDAMP	The visualization of differences between metabolite profiles acquired by hyphenated mass spectrometry techniques	http://mathdamp.iab.keio.ac.jp/
COMSPARI	Facilitate the analysis of “paired” samples	http://www.biomechanic.org/comspari/

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
2. Subramaniam S (1998) The Biology Workbench—a seamless database and analysis environment for the biologist. *Proteins* 32:1–2
3. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
4. Montpetit ML, Cassol S, Salas T, O'Shaughnessy MV (1992) OLIGSCAN: a computer program to assist in the design of PCR primers homologous to multiple DNA sequences. *J Virol Methods* 36:119–128

Chapter 16

Molecular Phylogenetics of Microbes

Surajit Das and Hirak Ranjan Dash

Abstract

Molecular phylogenetics or molecular systematic is the use of molecular structure to gain information on an organism's evolutionary relationships which is expressed as a phylogenetic tree. The impact of molecular systematic on bacterial classification has been profound. In the year 1977, Woesean revolution occurred when Carl Woese, a chemist working on relative isolation compared 16S rRNA sequences to study the classification of microorganisms. This molecular approach revealed three (Archea, Bacteria, Eukarya), rather than five (Animalia, Plantae, Fungi, Monera, Protists) primary divisions of life to describe extraordinary levels of microbial diversity. Through cells different apparatus provides numerous information related to an organism but SSU rDNAs (genes coding for small subunit ribosomal RNA) offer a quality and quantity of information which make them one of the most useful macromolecular descriptors of microorganisms. However, 16S rRNA sequence analysis has been criticized as some of the cases of lateral gene transfer reported in those genes. To avoid this ambiguity along with the analysis of 16S rRNA gene, certain housekeeping genes like *gapA*, *groEL*, *gyrA*, *ompA*, *pgi* are also recommended. The targeted genes can be analyzed in various ways either by documenting the gel in which the amplified products run or by analyzing the sequences of the targeted genes of interests. This chapter will focus on various molecular techniques involving gel-based techniques, sequence-based techniques, analyzing Dendrogram, and Cladograms.

16.1 Introduction

Molecular phylogenetics or molecular systematic is the use of molecular tools to gain information on an organism's evolutionary relationships which is expressed as a phylogenetic tree. The impact of molecular systematic on bacterial classification has been profound. In the year 1977, Woesean revolution occurred when Carl Woese, a chemist working on relative isolation compared 16S rRNA sequences to study the classification of microorganisms. This molecular approach revealed three (Archea, Bacteria, Eukarya), rather than five (Animalia, Plantae, Fungi, Monera, Protists) primary divisions of life to describe extraordinary levels of microbial diversity. Though cells different apparatus provides

numerous information related to an organism but SSU rDNAs (genes coding for small subunit ribosomal RNA) offer a quality and quantity of information which make them one of the most useful macromolecular descriptors of microorganisms. SSU rDNAs are widely used as informative biomarkers as:

- They are essential components of protein synthesis machinery and therefore are ubiquitously distributed and functionally conserved in all organisms.
- They lack the interspecies horizontal gene transfer found with many prokaryotic genes.
- They are readily isolated and identified.
- They contain diagnostic variable regions interspersed among highly conserved regions of primary and secondary structures, permitting phylogenetic comparisons to be inferred over a broad range of evolutionary distance.

However, 16S rRNA sequence analysis has been criticized as some of the cases of lateral gene transfer reported in those genes [1]. To avoid this ambiguity along with the analysis of 16S rRNA gene, certain housekeeping genes like *gapA*, *groEL*, *gyrA*, *ompA*, *pgi* are also recommended. The targeted genes can be analyzed in various ways either by documenting the gel in which the amplified products run or by analyzing the sequences of the targeted genes of interests.

Phylogenetic analysis is also known as molecular taxonomy. It uses the representation of evolutionary information in the form of phylogenetic trees. There are several methods for constructing phylogenetic tree. One of the most popular tools is PHYLIP (PHYLogeny Inference Package). The tree of life (<http://evolution.genetics.washington.edu/phylip.html>) is a collaborative Internet project containing information about phylogeny and biodiversity.

Certain softwares come in to play to draw the dendrograms to show the evolutionary history of the organism based upon the sequence variation of the gene of interest. Thus, to study molecular phylogeny of an organism, certain techniques like gel-based and sequence-based techniques are used followed by the drawing of dendrogram and cladogram to know the phylogeny and relatedness among organisms.

The genetic fingerprinting technique called denaturing gradient gel electrophoresis (DGGE) of PCR-amplified ribosomal DNA for microbial typing first came to light in the year 1993 [2]. In DGGE, DNA fragments of the same length but with different sequences can be separated. The polymerase chain reaction of environmental DNA can generate templates of differing DNA sequence that represent many of the dominant microbial organisms. However, since PCR products from a given reaction

are of similar size (bp), conventional separation by agarose gel electrophoresis results only in a single DNA band that is largely non-descriptive. DGGE can overcome this limitation by separating PCR products based on sequence differences that results in differential denaturing characteristics of the DNA. Once generated, fingerprints can be uploaded into databases in which fingerprint similarity can be assessed to determine microbial structural differences between environments or among treatments. Furthermore, with the breadth of PCR primers available, DGGE can also be used to investigate broad phylogenies or specific target organisms such as pathogens or xenobiotics degraders. In case of DGGE the separation is based on the decreased electrophoretic mobility of a partially melted double stranded DNA molecule in polyacrylamide gels containing a linear gradient of DNA denaturants (a mixture of urea and formamide). The percentage of GC content of the sequence plays a role during separation. As the GC bond pair is stronger than that of AT bond pair, the sequence with high GC content resists to denaturation and travels a longer distance where as the sequence with lower GC content denatures faster and cannot migrate for longer distance. As a result of which a specific banding pattern is observed for a same amplified product with sequence variation. It is highly useful for the estimation of microbial diversity of an environment by metagenomic approach.

Amplified rDNA restriction analysis is the extension of the technique RFLP to the gene encoding small ribosomal subunit of bacteria. The technique was originally developed by [3] to characterize *Mycobacterium* species. However, the method has been used by many more for characterization of other bacterial species also. Analysis of the patterns is done with methods used for RAPD patterns. Clusters of related bacteria can be represented in the form of a cardiogram or hologram. The data can be subsequently used for generating a phylogram or cladogram which can be used to plot a phylogenetic tree that would indicate the relationship of the organisms based on the restriction pattern obtained from their respective 16S genes. It is assumed that the related organisms give same restriction pattern by digestion with a particular restriction enzyme. The frequency of a random occurrence of a restriction site, a 4 bp sequence can occur as once in 256 bp repeats. When a tetra cutter restriction enzyme is applied along with 16S amplified sequence, it digests the sequence at particular sites to give a particular restriction pattern which is known to be the signature of the organism. The restriction pattern of one organism can be compared with the other to know the phylogenetic relationship between them. However for the sake of statistical significance, at least three restriction enzymes should be used to overcome the probability of certain restriction enzymes to yield similar patterns for unrelated organisms.

Restriction fragment length polymorphism (RFLP) is a generalized technique applied for phylogeny and fingerprinting analysis of bacteria based on the use of restriction enzymes which cut DNA at specific sequences. It is assumed that genetic related strains have similar distribution of restriction sites in their genome. But there are certain disadvantages like pattern complexity occur during RFLP as there is generation of too many number of fragments when any common 4–6 cutter restriction enzyme cut the genomic DNA. In order to avoid this problem, PCR RFLP should be performed for bacterial strains with particular characteristics like heavy metal resistance, antibiotic resistance, etc.

Polymerase chain reaction-based analysis of 16S rRNA genes is a powerful and essential tool for studies of bacterial diversity, community structure, evolution, and taxonomy. It enables us to detect and identify cultivable bacteria, as-yet uncultivable bacteria and in recent years it has led to an enormous increase in our knowledge of bacterial taxonomy. The rRNA is the most conserved gene of all cells. Portions of the rDNA sequences from distantly related organisms are remarkably similar. Hence that sequences from distantly related organisms can be precisely aligned and the differences can be truly measured. Hence genes that encode the rRNA are used extensively to determine taxonomy, phylogeny, and to estimate rates of species divergence among bacteria. Thus the comparison of 16S rDNA sequences can show evolutionary relatedness among microorganisms. This work has been pioneered by Carl Woese.

rpoB is a bacterial housekeeping gene that code for part of an enzyme which synthesizes RNA i.e., it is the β subunit of bacterial RNA polymerase. $Rpo\beta$ is a highly conserved enzyme contained by many bacteria. The complete *rpoB* gene sequence length varies from 3,452 to 3,845 bp which exhibits interspecies homology and intraspecies divergence [4]. Thus it can be used as an alternative tool for identification and subsequently for the analysis of phylogeny. Though nowadays the works clearly illustrate the usefulness of these sequences for enterobacteriaceae family but it can be applied to almost all types of organisms including bacteria and archaea [5]. Though a house keeping gene *rpoB* is also prone to mutations in case of bacteria. Distinct nucleotide substitutions in the sequences can lead a bacterium to confer resistance towards rifampin [6]. *rpoB*, the gene encoding the highly conserved subunit of the bacterial RNA polymerase, has been demonstrated to be a suitable target based on which the identification of enteric bacteria, spirochetes, bartonellas, and rickettsias are done [7]. The gene has been shown to be more discriminative than the 16S ribosomal DNA (rDNA).

gyrA is another bacterial housekeeping gene that codes for A subunit of DNA gyrase i.e., type II Topoisomerase. This is a gene of importance as it is of a 5–9 kb region which includes part of

an upstream *recF* gene, the whole *gyrB* and *gyrA* and about 1 kb of unknown downstream sequences. Transcription of *gyrA* gene increases in response to DNA relaxation. The most important point in relation to *gyrA* is that, resistance to quinolones in microorganisms is related to the acquisition of point mutation in the sequence of the QRDR of the *gyrA* for most microorganisms. In these microorganisms, the mutations in the gene encoding the subunit A of the DNA gyrase can confer quinolone resistance and overexpression of efflux pump may play a complementary role in quinolone resistance acquisition [8].

The differences in QRDR sequences of the *gyrA* gene allow differentiation of species by PCR-Restriction Fragment Length Polymorphism analysis using NcoI restriction enzyme [9]. This enzyme has different restriction sites in the sequence of *gyrA* gene depending upon the fact that to which organism the *gyrA* gene belongs to. However, the level of quinolone resistance seems to depend upon the type of amino acid substitution and on the amino acid substituted. Keeping in mind the selective variations in the sequences of *gyrA* gene the sequences of this gene can be compared with each other to draw the phylogenetic tree among them. However, this sequence-based housekeeping gene has been restricted to most pathogenic strains having characteristics features of quinolone resistance and mostly applicable for *Staphylococcus*, *Streptococcus*, *Helicobacter*, and *Corynebacterium* [10].

The sequencing of DNA and proteins has become easy and fast with the use of automated tools. When a set of sequences are present, the evolutionary relationships among genes and among organisms can be constructed. A phylogeny illustrates the relationships between the sequences. Analysis of phylogeny of a family (of related nucleic acid or protein sequences) is done by determination of how the family might have been derived during evolution. Molecular evolution is based on mutations impacting the DNA to change. Mutations can occur when there are errors in DNA replication or repair. There are various models of DNA change. The selection of model is one of the fundamental decisions to conduct the phylogenetic analysis.

The number or types of changes in the residues of a Multilocus Sequence Alignment (MSA) can be used to start a phylogenetic analysis. Each column in the MSA denotes mutations that occur at one site during the evolution of the sequence family. This information can be used to evaluate the positions in the sequences which are conserved and which diverge from a common ancestor sequence.

Evolutionary relationships can be represented using phylogenetic trees. A tree is a 2D graph showing evolutionary relationships among organisms. The tree is composed of nodes (a point where branches bifurcate) representing the taxa and branches representing the relationships among the taxa. The lengths of the branches are often drawn proportional to the number of sequence changes in the

branch and hence can represent the divergence. Two sequences that are very much alike will be located as neighboring outside branches and would be joined to a common branch beneath them. Usually phylogenetic analysis methods assume that each position in the protein or nucleic acid sequence changes independently of others. A clade is a group of organisms whose members share homologous features derived from a common ancestor. There are various programs available for performing various phylogenetic operations. Different programs and program options are different for DNA and protein sequences. Some of the most popular are *PlyloBLAST*, *Phylip*, *PAUP*, *PAML*, *Clustal W*, *ClustalX*, etc.

In a dendrogram, the trees are constructed by similarities of sequences which do not necessarily reflect evolutionary relationships. Distance methods also called phenetic methods compress all of the individual differences between pairs of sequences into a single number. Character-based methods are also called cladistic methods. The trees are calculated by considering the various possible pathways of evolution and are based on parsimony or likelihood methods. The resulting tree is called a cladogram. Cladistic methods use each alignment position as evolutionary information to build a tree.

There are two major types of cladistic methods based on Parsimony and based on Maximum likelihood. In the parsimony methods, for each position in the alignment, all possible trees are evaluated and are given a score based on the number of evolutionary changes needed to produce the observed sequence changes. The most parsimonious tree is the one with the fewest evolutionary changes for all sequences to derive from a common ancestor. This is a more time-consuming method than the distance methods. The maximum likelihood method also uses each position in an alignment. It evaluates all possible trees and calculates the likelihood for each tree using an explicit model of evolution. The likelihoods for each aligned position are then multiplied to provide the composite likelihood for each tree. The tree with the maximum likelihood is the most possible tree. This is the slowest method of all but may give the best result and the most information about the tree. There are various computer based softwares for such analysis like *PHYLIP*, *DNAPARS* and *fastDNAmI*.

16.2 Materials

16.2.1. DGGE

- Bacterial genomic DNA
- 16S rDNA forward primers (GC331F) [CGC CCG CGC GCG GCG GGC GGG GCG GGG GCG CGG GGG GTC CTA CGG GAG GCA GCA GT]

- 16S rDNA reverse primers (797R) [GGA CTA CCA GGG TAT CTA ATC CTG TT]
- dNTPs
- Taq DNA polymerase
- PCR reaction buffer
- PCR tubes and tips

Materials required for gradient polyacrylamide gel electrophoresis.

16.2.2. ARDRA

- Genomic DNA
- 16S forward [AGA GTT TGA TCC TGG CTC AG] primer
- 16S reverse [ACG GCT ACC TTG TTA CGA CTT] primer
- dNTPs
- Taq DNA polymerase
- PCR reaction buffer
- PCR tubes and tips
- Restriction enzymes (tetra cutters)
- Gel apparatus and reagents

16.2.3. 16S rRNA Gene Sequencing

- Genomic DNA
- PCR buffer
- Magnesium chloride
- dNTPs
- Universal forward and reverse primers
- Taq DNA polymerase
- Automatic DNA sequencer
- NCBI database

16.2.4. *rpoB* Gene Sequencing

- Genomic DNA
- PCR buffer
- Magnesium chloride
- dNTPs
- *rpoB* forward [2643F: 5'CAA TTC ATG GAC CAA GC 3'] primer
- *rpo B* reverse [3241R: 5' GCI ACI TGI TCC ATA CCT GT 3'] primer
- Taq DNA polymerase
- DNA sequencer
- *rpoB* database

16.2.5. *gyrA* Gene Sequencing

- Genomic DNA
- PCR buffer
- Magnesium chloride
- dNTPs
- *gyrA* Forward [5' GCG GCT ACG TAA AGT CC] primer
- *gyrA* Reverse [5' CCG CCG GAG CCG TTC AT 3'] primer
- Taq DNA polymerase
- DNA sequencer
- *gyrA* Database

16.3 Methods

16.3.1. DGGE

16.3.1.1. PCR

Reaction mixture	Cycling conditions	
10× buffer-10 µl	Denaturation-94 °C for 5 min	
10 mM dNTPs mixture-2.5 µl	Denaturation-94 °C for 30 s	} 30 Cycles
5 µM forward primer-10 µl	Annealing-55 °C for 30 s	
5 µM reverse primer-10 µl	Extension-72 °C for 2 min	
2.5 U/µl of Taq polymerase-1 µl	Extension-72 °C for 5 min	
Water-39 µl	Holding-4 °C forever	
Template DNA-2 µl		

16.3.1.2. Sample Preparation

- Load 50 µl of the amplified DNA per well.
- Prior to loading add equal volume of 2× gel loading dye to the sample.

16.3.1.3. Preheating of the Loading Buffer

- Fill the electrophoresis tank with 7 L of 1× TAE running buffer.
- Set the temperature to 60 °C and the temperature ramp rate to 200 °C/h.

16.3.1.4. Casting

Denaturing Gradient Gel

- Prepare low density solution and high density solution.
- Add the 0.09 % (v/v) each of ammonium persulfate and TEMED solutions. Mix by inverting several times.
- Draw both high and low density solutions to respective syringes; care should be taken so that there will be no air bubble in the syringes.
- Place the syringes in to the gradient delivery system syringe holder. Rotate the wheel clockwise slowly to deliver the gel solution.
- Carefully insert the comb to the desired well depth. It will take about 30–60 min to polymerize the gel.
- After polymerization remove the comb by pulling it straight up slowly and gently.
- Load the sample DNA prepared earlier and continue electrophoresis by assembling the electrophoresis apparatus.

16.3.1.5. Running the Gel

- Attach the electrical leads to a suitable DC power supply supplied by the system.
- Run the gel at 130 V for overnight.

16.3.1.6. Staining

and Documenting the Gel

- Remove the gel from the glass plate.
- Place the gel in to a dish containing 250 ml of running buffer and 25 μ l of 10 mg/ml ethidium bromide. Stain for 5–15 min.
- After staining carefully transfer the gel into a dish containing 250 ml of running buffer. Destain for 5–20 min.
- Photograph the system in Gel documentation system.

16.3.1.7. DGGE Banding

Patterns and Statistical Analysis

The digitized DGGE images were analyzed using Gel Pro (Bio-Rad). We combined four images of a DGGE gel using Adobe Photoshop CS4 (Adobe Systems Inc., USA) to amend the relative mobility between different gels. We developed a binary matrix containing data on the presence/absence of bands and a proportional matrix displaying the percentage of each band based on relative pixel intensities for each lane.

The Shannon–Weaver index (H_p) was calculated using the function $H_p = -\sum P_i \ln P_i$, where P_i is the importance probability of the bands in a gel lane [11] and $P_i = n_i/N$, where n_i is the intensity of the individual bands and N is the sum of the intensities for all bands in a lane. We generated an unweighted pair-group method with an arithmetic mean (UPGMA) tree using Ntsys, based on the band intensity matrix, to examine the similarities in

the DGGE profiles among all 40 samples, [12]. Correlations between the DGGE profiles and sediment physicochemical properties were analyzed by CCA using Canoco (Microcomputer Power, USA) [13]. The statistical significance of the relationship was assessed by Monte Carlo analysis using 1,000 permutations.

16.3.1.8. *Phylogenetic Analysis of the Sequenced DGGE Bands*

Analyze the fingerprinting pattern and determine the phylogenetic relationship among the tested bacterial strains. Then construct neighbor-joining [14] tree using MEGA 4.1 [15]. The bootstrap values were replicated 1,000 times.

16.3.2. **ARDRA**

16.3.2.1. *PCR for Amplification of 16S rDNA Sequence*

Reaction mixture	Cycling conditions	
10× buffer-5 µl	Denaturation-94 °C for 5 min	
10 mM dNTPs mixture-1 µl	Denaturation-94 °C for 30 s	} 30 Cycles
10 µM forward primer-1 µl	Annealing-55 °C for 30 s	
10 µM reverse primer-1 µl	Extension-72 °C for 2 min	
2.5 U/µl of Taq polymerase-1 µl	Extension-72 °C for 5 min	
Water-39 µl	Holding-4 °C forever	
Template DNA-2 µl		

16.3.2.2. *Restriction Digestion*

- 10 µl of 16S rDNA amplified product was taken in a micro centrifuge tube.
- It was digested with 5 U of restriction enzyme AluI and restriction buffer is added to make a total volume of 20 µl.
- It was incubated for 2 h at 37 °C for restriction digestion.

16.3.2.3. *Electrophoresis*

- Run the restriction digested product on 3 % agarose at 130 V for 3 h.
- Analyze the banding pattern with Gel Documentation System and determine the phylogenetic relationship.

16.3.3. 16S rRNA Gene Sequencing

16.3.3.1. 16S rRNA Gene Amplification

Reaction mixture	Cycling conditions	
10× buffer-5 µl	Denaturation-94 °C for 5 min	
10 mM dNTPs mixture-1 µl	Denaturation-94 °C for 30 s	} 30 Cycles
10 µM forward primer- 1 µl	Annealing-55 °C for 30 s	
10 µM reverse primer-1 µl	Extension-72 °C for 2 min	
2.5 U/µl of Taq polymerase-1 µl	Extension-72 °C for 5 min	
Water-39 µl	Holding-4 °C forever	
Template DNA-2 µl		

16.3.3.2. Purification of PCR Product

Purify the PCR product either by readily available gel elution kits or by PCR purification kits.

16.3.3.3. Sequencing

The purified PCR product is sequenced either by dideoxy chain termination method or chemical method or by pyrosequencing.

16.3.3.4. BLAST

The sequences obtained can be compared with the databases like NCBI, expasy to know the percentage of similarity basing upon which the phylogenetic relationship of the organism with others can be determined.

16.3.4. rpoB Gene Sequencing

16.3.4.1. rpoB Gene Amplification

Reaction mixture	Cycling conditions	
10× buffer-5 µl	Denaturation-94 °C for 5 min	
10 mM dNTPs mixture-1 µl	Denaturation-94 °C for 30 s	} 30 Cycles
10 µM forward primer-1 µl	Annealing-5 °C for 30 s	
10 µM reverse primer-1 µl	Extension-72 °C for 2 min	
2.5 U/µl of Taq polymerase-1 µl	Extension-72 °C for 5 min	

(continued)

Reaction mixture	Cycling conditions
Water-39 µl	Holding-4 °C forever
Template DNA-2 µl	

16.3.4.2. Purification of PCR Product Purify the PCR product either by readily available gel elution kits or by PCR purification kits.

16.3.4.3. Sequencing The purified PCR product is sequenced either by dideoxy chain termination method or chemical method or by pyrosequencing.

16.3.4.4. BLAST The sequences obtained can be compared with the databases like NCBI, ExPASy to know the percentage of similarity basing upon which the phylogenetic relationship of the organism with others can be determined.

16.3.5. *gyrA* Gene Sequencing

*16.3.5.1. **gyrA** Gene Amplification*

Reaction mixture	Cycling conditions	
10× buffer-5 µl	Denaturation-94 °C for 5 min	
10 mM dNTPs mixture-1 µl	Denaturation-94 °C for 30 s	} 30 Cycles
10 µM forward primer-1 µl	Annealing-55 °C for 30 s	
10 µM reverse primer-1 µl	Extension-72 °C for 2 min	
2.5 U/µl of Taq polymerase-1 µl	Extension-72 °C for 5 min	
Water-39 µl	Holding-4 °C forever	
Template DNA-2 µl		

16.3.5.2. Purification of PCR Product Purify the PCR product either by readily available gel elution kits or by PCR purification kits Gen elute[®], Sigma–Aldrich.

16.3.5.3. Sequencing The purified PCR product is sequenced either by dideoxy chain termination method or chemical method or by pyrosequencing by using automated DNA sequencers.

16.3.5.4. BLAST BLAST (Basic Local Alignment Search Tool) is a similarity search program developed at NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>). It is available as a free service over the internet that provides very fast, accurate and sensitive database searching.

Table 16.1
URL address of some commonly used databases

Name of the program	Address (URL)
BLAST Network Service on ExPASy	http://us.expasy.org/tools/blast/
BLAST at EMBnet-CH/SIB (Switzerland)	http://www.ch.embnet.org/software/Bottom BLAST.html?
BLAST at NCBI	http://www.ncbi.nlm.nih.gov/BLAST/
WU-BLAST at the EBI	http://www.ebi.ac.uk/blast2/
BLAST at PBIL (Lyon)	http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_blast.html

BLAST uses a heuristic algorithm that seeks local as opposed to global alignments and is therefore able to detect relationships among sequences that share only isolated regions of similarity. The sequences obtained can be compared with the databases like NCBI, ExPASy to know the percentage of similarity basing upon which the phylogenetic relationship of the organism with others can be determined. Some of the most used databases are as follows (Table 16.1):

16.3.6. Clustal X

16.3.6.1. Obtaining a Related Sequence by a BLAST Search

We are already having a particular nucleic acid and protein sequence of our interest and we need to find other sequences that are related to it, i.e., another sequence sufficiently similar to the sequence of interest so that we believe, the two sequences share a common ancestor. This may be done by the following steps:

- Go to <http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast?Jform=0>
- Open the electronic file containing your sequence of interest and copy the sequence.
- Return to the BLAST page and click the text box below where it shows **Sequence in FASTA** format.
- In this configuration type the accession number for the sequence of interest into the text box.
- Click the **Submit Query** button to submit the sequence to BLAST.
- At the left of each sequence is the name and accession number of the file, which is in one of the databases that has been searched, in the middle is a brief description of the file.

- **Bit Score**—For DNA sequences, the bit score is calculated by assigning a score of 1 to each match and 0 to each mismatch, then subtracting penalties for gaps. The higher the bit score, the more closely related that sequence to the sequence of interest.
- At the far right the number is the E value, which is the number of such matches to the current non-redundant sequences database that are expected by chance alone. The smaller the E value, the more likely that the similarity is real (This is a critical choice because that list will be the set of sequences from which we construct our phylogenetic tree).
- Downloading the selected sequence, click the **FASTA** button and the same file will be saved in the FASTA format.

16.3.6.2. Creating Multiple Sequence Alignment

A pair of sequences can be aligned by writing one above the other in such a way as to maximize the number of residues (nucleotides or aminoacids) that match by introducing gaps (spaces) into one or other sequences. Biologically those gaps are assumed to represent insertions or deletions that occurred as the sequences diverged from a common ancestor. A scoring system is used so that matching residues get some sort of positive numerical score, and gaps get sort of negative score, or gap penalty. An alignment program seeks an arrangement that maximizes the net score. Gap penalties are typically set by the user. ClustalX is an updated version of clustalW which is one of the best tools for creating multiple alignments. Follows the steps for alignment:

- Create your input file in FASTA format
- Copy the entire sequence file, including the first line, and paste it in to the CelF. in file. In, clustal treats everything between ">" and the first space as the sequence name
- Save the file in plain text or ASCII format, this is important because ClustalX will not recognize word or any other word processor file
- Go to http://www.biozentrum.unibas.ch/biophit/clustal/ClustalX_help.html
- Pull down the File menu and choose the Load sequence menu item
- Pull the Alignment menu down to choose the Alignment parameters (pairwise alignments: slow to accurate, Gap opening Penalty 15.00 and Gap Extension Penalty to 0.30, Delay Divergent sequences 25 %, Output format options to Nexus and change the Clustal Sequence Numbers on)
- Choose do complete alignment under the alignment menu
- After alignment is over Eliminate truncated sequences

- Delete nonhomologous regions from the sequences
- Change the gap penalties
- Use PAUP to create a tree

16.3.6.3. Creating Tree

The phylogenetic tree can be drawn by the following steps:

- Open <http://paup.csit.fsu.edu/>
- Create the input file in Nexus format
- Pull down the Analysis menu and choose distance
- Using same Analysis menu, choose Neighbor joining/UPGMA
- Accept the computer generated number and click OK
- Click the Preview button and you can see a slanted phylogenetic tree which can be printed using print tree.

16.3.7. Cladogram

Obtain nexus file by multiple sequence alignment and follow the steps to draw cladogram:

- Open PAUP link
- Pull down the Alignment menu, be sure that Parsimony is checked
- Choose Heuristic Search
- In the resulting dialogue leave everything in its default state
- Click the Search button

When search is completed it will show a close button and will indicate the number of trees that were created. The trees are now in memory and it can be printed by selecting print trees from the Trees menu or can be view by TreeView (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>) software.

References

1. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM (1999) Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399:323–329
2. Muyzer G, De Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 59:695–700
3. Vaneechoutte M, Riegel P, Briel D, Monteil H, Verschraegen G, Rouck A, Claeys G (1995) Evaluation of the applicability of amplified rDNA restriction analysis (ARDRA) to identification of the genus *Corynebacterium*. *Res Microbiol* 146:633–641
4. Regnery RL, Spuill CL, Plikaytis BD (1991) Genotypic identification of rickettsiae and estimation of intraspecies sequence divergence for portions of two rickettsial genes. *J Bacteriol* 173:1576–1589
5. Mollet C, Drancourt M, Raoult D (1997) *rpoB* sequence analysis as a novel basis for bacterial identification. *Mol Microbiol* 26:1005–1011

6. Hirano K, Abe C, Takahashi M (1999) Mutations in the *rpoB* gene of rifampin-resistant *Mycobacterium tuberculosis* strains isolated mostly in asian countries and their rapid detection by line probe assay. *J Clin Microbiol* 37:2663–2666
7. Drancourt M, Raoult D (2002) *rpoB* gene sequence based identification of *Staphylococcus* species. *J Clin Microbiol* 40:1333–1338
8. Hooper DC (2001) Mechanisms of action of antimicrobials: focus on fluoroquinolones. *Clin Infect Dis* 32:9–15
9. Bachoual R, Ouabdesselam S, Mory F, Lascons C, Soussy CJ, Tankovic J (2001) Single and double mutational alterations of *gyrA* associated with fluoroquinolone resistance in *Campylobacter jejuni* and *Campylobacter coli*. *Microb Drug Resist* 3:257–261
10. Sierra MJ, Martinez LM, Vazquez F, Giralt E, Vila J (2005) Relationship between mutations in the *gyrA* gene and quinolone resistance in clinical isolates of *Corynebacterium striatum* and *Corynebacterium amycoiatum*. *Antimicrob Agents Chemother* 49:1714–1719
11. Eichner CA, Erb RW, Timmis KN, Dobler IW (1999) Thermal gradient gel electrophoresis analysis of bioprotection from pollutant shocks in the activated sludge microbial community. *Appl Environ Microbiol* 65:102–109
12. Rohlf FJ (1987) NTSYS-pc: microcomputer programs for numerical taxonomy and multivariate analysis. *Am Stat* 41:330
13. Braak CJF, Verdonschot PFM (1995) Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sci* 57:255–289
14. Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
15. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9:299–306

Chapter 17

Microarray Technology: Basic Concept, Protocols, and Applications

P.P. Dubey and Dinesh Kumar

Abstract

Microarray is one such technology which enables the researchers to investigate and address issues which were once thought to be nontraceable. Over the past few years, this powerful technology has been used to explore transcriptional profiles and genome differences for a variety of microorganisms, greatly facilitating our understanding of microbial metabolism. With the increasing availability of complete microbial genomes, DNA microarrays are becoming a common tool in many areas of microbial research, including microbial physiology, pathogenesis, epidemiology, ecology, phylogeny, and pathway engineering. One can analyze the expression of many genes in a single reaction quickly and in an efficient manner. DNA Microarray technology helps in the identification of new genes and in knowing about their functioning and expression levels under different conditions, comparative genomics, and SNPs identification. This chapter has outlined the principle of microarray technology, types of microarray, their basic protocols, and applications.

17.1 Introduction

Microarray technology evolved from Southern blotting, where fragmented DNA is attached to a substrate and then probed with a known gene or fragment. The use of a collection of distinct DNAs in arrays for expression profiling was first described in 1987, and the arrayed DNAs were used to identify genes whose expression is modulated by interferon. These early gene arrays were made by spotting cDNAs onto filter paper with a pin-spotting device. The use of miniaturized microarrays for gene expression profiling was first reported in 1995, and a complete eukaryotic genome (*Saccharomyces cerevisiae*) on a microarray was published in 1997. It is known that thousand of genes and their products (i.e., RNA and Proteins) in a given living organism function in a complex and orchestrated way that create the mystery of life. However, traditional methods in molecular biology generally work on one gene in one experiment basis, which means that the throughput is very limited and it is very hard to know the

complete picture of gene function in an organism. The invention of polymerase chain reaction (PCR) produced a surge in new experiments. Because PCR accelerated and simplified the procedures previously performed much more laboriously by traditional molecular cloning, it quickly found use in experimental molecular biology. In the past several years, a new technology, called DNA microarray, has attracted tremendous interest among biologists. DNA microarrays are a powerful tool for the investigation of various aspects of prokaryotic biology because they allow the simultaneous monitoring of the expression of all genes in any bacterium. They offer a more holistic approach to study cellular physiology and therefore complement the traditional “gene-by-gene” approaches [1]. The term DNA microarray was coined in publications from the laboratory of DeRisi [2] and Schena [3]. This technology promises to monitor the whole genome in single chip so that the researcher can have a better picture of the interactions among thousands of genes simultaneously. Molecular biology research evolves through the development of the technologies used for carrying them out. It is not possible to research on a large number of genes using traditional methods. DNA microarray technology has empowered the scientific community to understand the fundamental aspects underlining the growth and development of life as well as to explore the genetic causes of anomalies occurring in the functioning of the human body.

17.1.1. Principle

A typical microarray experiment involves the hybridization of an mRNA molecule to the DNA template from which it is originated. Many DNA samples are used to construct an array. The amount of mRNA bound to each site on the array indicates the expression level of the various genes. This number may run in thousands. All the data are collected and a profile is generated for gene expression in the cell. Additional advantages of microarrays are that they are highly sensitive and are small. The microarray constitutes a large array of highly ordered immobilized target sequences attached to a solid surface. Each target sequence corresponds to a different gene; mRNA is taken from a particular cell line or tissue combined with some sort of marker to generate a labeled sample. This sample is then hybridized onto the target sequences of the microarray. The marked sample will bind with its complementary sequence so for each gene the amount of marker is detected and this provides a level of expression for that gene. Due to the large number of measurements that are taken simultaneously this produces a huge amount of data. Various techniques have been developed to constructively deal with the size and variability of microarray data.

An array is an orderly arrangement of samples where matching of known and unknown DNA samples is done based on base pairing rules. An array experiment makes use of common assay systems such as microplates or standard blotting membranes.

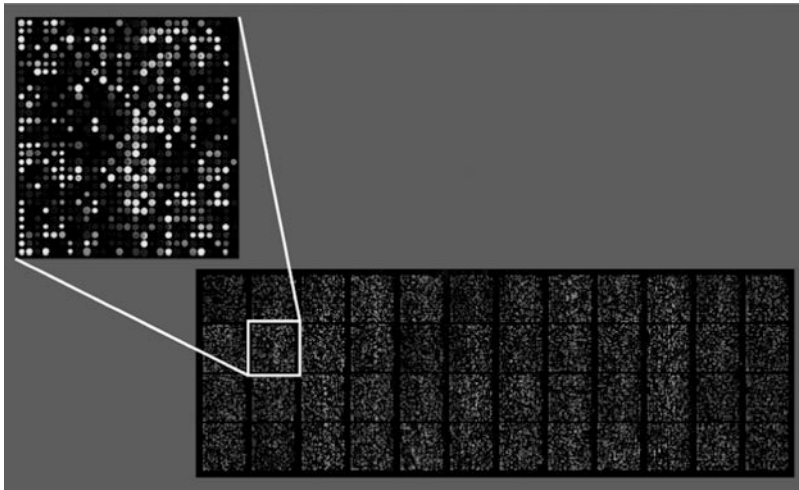


Fig. 17.1 Example of an approximately 40,000 probe spotted Oligo microarray.

The sample spot sizes are typically less than 200 μm in diameter and usually contain thousands of spots (Fig. 17.1). Thousands of spotted samples known as probes (with known identity) are immobilized on a solid support (microscope glass slides or silicon chips or nylon membrane). The spots can be DNA, cDNA, or oligonucleotides. These are used to determine complementary binding of the unknown sequences, thus allowing parallel analysis for gene expression and gene discovery. An experiment with a single DNA chip can provide information on thousands of genes simultaneously. An orderly arrangement of the probes on the support is important as the location of each spot on the array is used for the identification of a gene.

17.1.2. Types of Microarrays

Depending upon the kind of immobilized sample used to construct arrays and the information fetched, the Microarray experiments can be categorized in three ways:

- (a) **Microarray expression analysis:** In this experimental setup, the cDNA derived from the mRNA of known genes is immobilized. The sample has genes from both the normal and the diseased tissues. Spots with more intensity are obtained for diseased tissue gene if the gene is overexpressed in the diseased condition. This expression pattern is then compared to the expression pattern of a gene responsible for a disease.
- (b) **Microarray for mutation analysis:** For this analysis, the researchers use gDNA. The genes might differ from each other by as less as a single nucleotide base. A single base difference between two sequences is known as Single Nucleotide Polymorphism (SNP) and detecting them is known as SNP detection.

- (c) Comparative genomic hybridization: It is used for the identification in the increase or decrease of the important chromosomal fragments harboring genes involved in a disease.

According to the nature of the probe, microarray can be classified as (a) double-stranded DNA microarrays and (b) oligonucleotide DNA microarrays. There are two major types of probes that are used with DNA microarray printers: double-stranded DNA and oligonucleotides. Double-stranded DNA commonly results from PCR amplification [4]. A 200 to 800bp length of amplified DNA is recommended, but larger fragments of up to 1.3kb length also work [5]. In typical microarray design, each probe DNA corresponds to one gene. This represents the original type of DNA microarrays where cDNA molecules from *Arabidopsis thaliana* were amplified by PCR and spotted.

17.1.3. Affymetrix GeneChips

The most prominent microarrays with in situ synthesized probes are the GeneChips manufactured by Affymetrix, Santa Clara, CA, USA [6]. They are produced by chemical synthesis of the oligonucleotides directly on the coated quartz surface of the array [7]. This technology allows very high feature densities. It is typical to have 400,000 features on a commercial array [8]. Therefore, they are called high-density oligonucleotide arrays. GeneChips are produced in a unique photolithographic process analogous to the methods used for production of microelectronics chips in combination with chemical reactions developed for combinatorial chemistry. A quartz wafer is coated with a narrow layer of a light-sensitive compound. This coating prevents the covalent coupling of an activated nucleotide. Exposure to light causes the removal of the chemical protection groups from the surface. Subsequently applied reactive derivatives of single nucleotides can then be coupled. The attached nucleotides again carry a light-sensitive protection group that has to be removed by illumination before coupling the next nucleotide. Lithographic masks are used to block or transmit light onto specific features, thereby determining the order of nucleotide to be coupled to the growing oligonucleotides. In repeated cycles of masking, light exposure, and coupling, oligonucleotides of 25 residues' length are synthesized on the chip surface. As the specificity of a probe of 25 nucleotides may not be high enough, each probe ("match") is accompanied by a negative control with a single differing base in the middle of the probe termed mismatch probe. Performance of probe and mismatch probe can therefore be used to detect and eliminate cross-hybridization. Probe and mismatch probe are called a probe pair. Usually, 11–15 probe pairs, called a probe set, are used to represent a single gene. The very high feature density in this type of microarray enables the high number of controls.

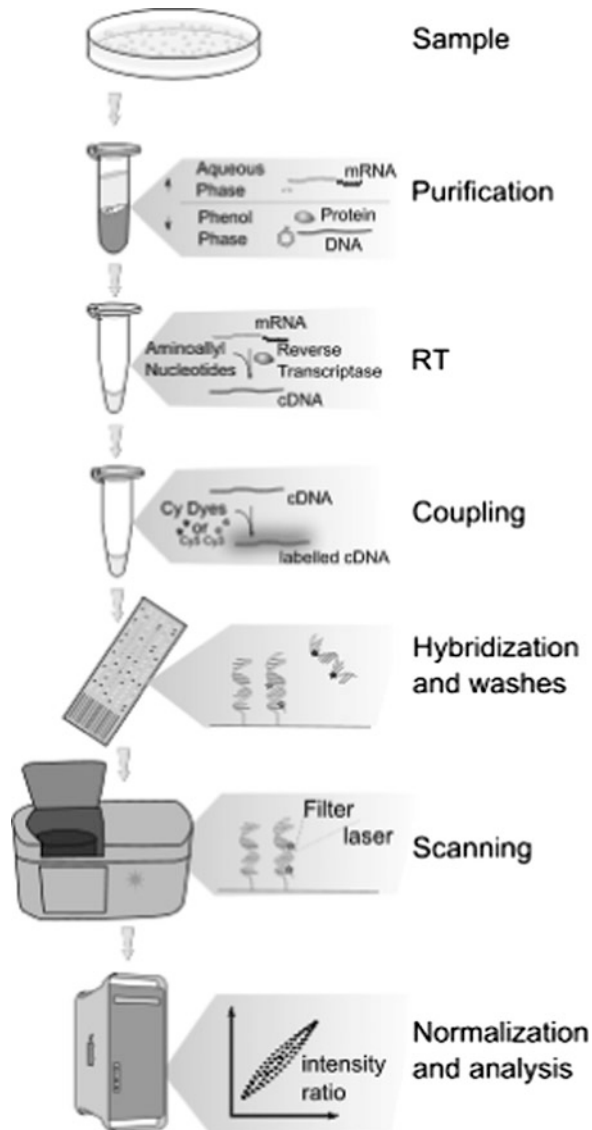


Fig. 17.2 Steps in microarray experiment.

17.1.4. Steps in Microarray Experiment

The following steps are involved in microarray experiments (Fig. 17.2):

1. The two samples to be compared (pairwise comparison) are grown/acquired, e.g., treated sample (case) and untreated sample (control).
2. The nucleic acid of interest is purified: this can be all RNA for expression profiling, DNA for comparative hybridization, or DNA/RNA bound to a particular protein which is

immunoprecipitated (ChIP-on-chip) for epigenetic or regulation studies. In this example total RNA is isolated (total as it is nuclear and cytoplasmic) by Guanidinium thiocyanate–phenol–chloroform extraction (e.g., Trizol) which isolates most RNA (whereas column methods have a cutoff of 200 nucleotides) and if done correctly has a better purity.

3. The purified RNA is analyzed for quality (by capillary electrophoresis) and quantity (by using a nanodrop spectrometer): if enough material ($>1\ \mu\text{g}$) is present the experiment can continue.
4. The labeled product is generated via reverse transcription and sometimes with an optional PCR amplification. The RNA is reverse transcribed with either poly-T primers, which amplify only mRNA, or random primers, which amplify all RNA which is mostly rRNA. In miRNA microarray an oligonucleotide is ligated to the purified small RNA (isolated with a fractionator) and then RT and amplified. The label is added either in the RT step or in an additional step after amplification if present. The sense that is labeled depends on the microarray, which means that if the label is added with the RT mix, the cDNA is on the template strand while the probes on the sense strand (unless they are negative controls). The label is typically fluorescent; only one machine uses radiolabels. The labeling can be direct (not used) or indirect which requires a coupling stage. The coupling stage can occur before hybridization (two-channel arrays) using aminoallyl-UTP and NHS amino-reactive dyes (like cyanine dyes) or after (single-channel arrays) using biotin and labeled streptavidin. The modified nucleotides (1 aaUTP: 4 TTP mix) are added enzymatically at a lower rate compared to normal nucleotides, typically resulting in 1 every 60 bases (measured with a spectrophotometer). The aaDNA is then purified with a column (using solution containing phosphate buffer as Tris contains amine groups). The aminoallyl group is an amine group on a long linker attached to the nucleobase, which reacts with a reactive dye. A dye flip is a type of replicate done to remove any Dye effects in two-channel dyes: one is labeled with cy3 and the other with cy5 and this is reversed in a different slide, e.g., in the presence of aminoallyl-UTP added in the RT mix.
5. The labeled samples are then mixed with a propriety hybridization solution which may contain SDS, SSC, dextran sulfate, a blocking agent (such as COT1 DNA, salmon sperm DNA, calf thymus DNA, PolyA, or PolyT), Denhardt's solution, and formamine.
6. This mix is denatured and added to a pin hole in a microarray, which can be a gene chip (holes in the back) or a glass

microarray which is bound by a cover, called a mixer containing two pinholes and sealed with the slide at the perimeter.

7. The holes are sealed and the microarray hybridized, either in a hyb oven, where the microarray is mixed by rotation, or in a mixer, where the microarray is mixed by alternating pressure at the pinholes.
8. After an overnight hybridization, all nonspecific binding is washed off (SDS and SSC).
9. The microarray is dried and scanned in a special machine where a laser excites the dye and a detector measures its emission.
10. The image is gridded with a template and the intensities of the features (several pixels make a feature) are quantified.
11. The raw data are normalized: the simplest way is to subtract the background intensity and then divide the intensities making either the total intensity of the features on each channel equal or the intensities of a reference gene and then the *t*-value for all the intensities is calculated. More sophisticated methods include *z*-ratio, loess and lowess regression, and RMA (robust multichip analysis) for Affymetrix chips (single-channel, silicon chip, in situ synthesized short oligonucleotides).

17.2 Typical Protocols and Major Steps for Microarray

17.2.1. Fabrication

This protocol describes the steps required to produce a cDNA microarray. Gene-specific DNA is produced by PCR amplification of purified template plasmid DNAs from cloned ESTs. The PCR product is purified by ethanol precipitation, thoroughly resuspended in $3\times$ SSC, and printed onto a poly-L-lysine-coated slide.

17.2.1.1. Slide Coating

Slides coated with poly-L-lysine have a surface that is both hydrophobic and positively charged. The hydrophobic character of the surface minimizes spreading of the printed spots, and the charge appears to help position the DNA on the surface in a way that makes cross-linking more efficient.

17.2.1.1.1. Materials, Reagents, and Solutions

- Gold seal microscope slides (#3011, Becton Dickinson, Franklin Lake, NJ)
- Ethanol (100 %)
- Poly-L-lysine (#P8920, Sigma, St. Louis, MO)

- 50-slide stainless steel rack, #900401, and 50-slide glass tank, #900401 (Wheaton Science Products, Millville, NJ)
- Sodium hydroxide
- Stir plate
- Stir bar
- Platform shaker
- 30-slide rack, #196, plastic, and 30 slide box, #195, plastic (Shandon Lipshaw, Pittsburgh, PA)
- Sodium chloride
- Potassium chloride
- Sodium phosphate dibasic heptahydrate
- Potassium phosphate monobasic
- Autoclave
- 0.2 mm Filter: nalgene
- Centrifuge: Sorvall Super 20
- Slide box (plastic with no paper or cork liners), (e.g., #60-6306-02, PGC Scientific, Gaithersburg, MD)
- 1 l glass beaker
- 1 l graduated cylinder

1 M Sodium Borate (pH 8.0)

- Dissolve 61.83 g of Boric acid in 900 ml of DEPC H₂O. Adjust the pH to 8.0 with 1 N NaOH. Bring volume up to 1 l. Sterilize with a 0.2 µm filter and store at room temperature.

Cleaning Solution

- H₂O 400 ml
- Ethanol 600 ml
- NaOH 100 g. Dissolve NaOH in H₂O. Add ethanol and stir until the solution clears. If the solution does not clear, add H₂O until it does.

Poly-L-Lysine Solution

- Poly-L-lysine (0.1 % w/v) 35 ml
- PBS 35 ml
- H₂O 280 ml

17.2.2. Procedure

1. Place slides into 50-slide racks and place racks in glass tanks with 500 ml of cleaning solution. Gold Seal Slides are highly recommended, as they have been found to have consistently low levels of autofluorescence. It is important to wear powder-

free gloves when handling the slides. Change gloves frequently, as random contact with skin and surfaces transfers grease to the gloves.

2. Place tanks on platform shaker for 2 h at 60 rpm.
3. Pour out cleaning solution and wash in H₂O for 3 min. Repeat wash four times.
4. Transfer slides to 30 slide plastic racks and place into small plastic boxes for coating.
5. Submerge slides in 200 ml poly-L-lysine solution per box.
6. Place slide boxes on platform shaker for 1 h at 60 rpm.
7. Rinse slides three times with H₂O.
8. Submerge slides in H₂O for 1 min.
9. Spin slides in centrifuge for 2 min at 400×g and dry slide boxes used for coating.
10. Place slides back into slide box used for coating and let stand overnight before transferring to new slide box for storage. This allows the coating to dry before handling.
11. Allow slides to age for 2 weeks on the bench, in a new slide box, before printing on them. The coating dries slowly, becoming more hydrophobic with time.

17.2.3. Slide Blocking

At the end of the print, slides are removed from the printer, labeled with the print identifier and the slide number by writing on the edge of the slide with a diamond scribe, and placed in a dust-free slide box to age for 1 week. It is useful to etch a line, which outlines the printed area of the slide, onto the first slide. This serves as a guide to locate the area after the slides have been processed, and the salt spots are washed off.

1. Place slides printed face up, in casserole dish and cover with cling wrap. Expose slides to a 450 mJ dose of ultraviolet irradiation in the Stratalinker. Slides should have been aged at ambient temperature in a closed slide box for 1 week prior to blocking.
2. Transfer slides to a 30-slide stainless steel rack and place rack into a small glass tank.
3. Dissolve 6.0 g succinic anhydride in 325 ml 1-methyl-2-pyrrolidinone in a glass beaker by stirring with a stir bar. Nitrile gloves should be worn and work carried out in a hemical fume hood while handling 1-methyl-2-pyrrolidinone (a teratogen).
4. Add 25 ml 1 M sodium borate buffer (pH 8.0) to the beaker. Allow the solution to mix for a few seconds, and then pour rapidly into glass tank with slides. Succinic anhydride hydrolyzes quite rapidly once the aqueous buffer solution is added.

To obtain quantitative passivation of the poly-L-lysine coating, it is critical that the reactive solution be brought in contact with the slides as quickly as possible.

5. Place the glass tank on a platform shaker in a fume hood for 20 min. Small particulates resulting from precipitation of reaction products will be visible in the fluid.
6. While the slides are incubating on the shaker, prepare a boiling H₂O bath to denature the DNA on the slides.
7. After the slides have incubated for 20 min, transfer them into the boiling H₂O bath. Immediately turn off the heating element after submerging the slides in the bath. Allow slides to stand in the H₂O bath for 2 min.
8. Transfer the slides into a glass tank filled with 100 % ethanol and incubate for 4 min.
9. Remove the slides and centrifuge at 400 rpm for 3 min in a horizontal microtiter plate rotor to dry the slides.
10. Transfer slides to a clean, dust-free slide box and let them stand overnight before hybridizing.

17.2.4. Printing

The variety of printers and pens for transferring PCR products from titer plates to slides precludes highly detailed descriptions of the process. The following steps provide a general description of the processing:

1. Pre-clean the print pens according to the manufacturer's specification.
2. Load the printer slide deck with poly-L-lysine-coated slides.
3. Thaw the plates containing the purified EST PCR products and centrifuge briefly, 2 min, at 1,000 rpm in a horizontal microtiter plate rotor to remove condensation and droplets from the seals before opening.
4. Transfer 5–10 μ l of the purified EST PCR products to a plate that will serve as the source of solution for the printer.
5. Printing with quill-type pens usually requires that the volume of fluid in the print source is sufficiently low and that when the pen is lowered to the bottom of the well, it is submerged in the solution to a depth of less than a millimeter. This keeps the pen from carrying a large amount of fluid on the outside of the pen shaft and producing variable, large spots on the first few slides printed.
6. Run a repetitive test print on the first slide. In this operation, the pens are loaded with the DNA solution, and then the pens serially deposit this solution on the first slide in the spotting pattern specified for the print. This test is run to check the size and shape of the specified spotting pattern, and its placement on

the slide. It also serves to verify that the pens are loading and spotting and that a single loading will produce as many spots as are required to deliver material to every slide in the printer.

7. If one or more of the pens is not performing at the desired level, re-clean or substitute another pen and test again.
8. If all pens are performing, carry out the full print.

17.2.5. Sample Labeling

1. If using an anchored oligo dT primer, anneal the primer to the RNA in the following 17 μ l reaction (use a 0.2 ml thin wall PCR tube so that incubations can be carried out in a PCR cyclor).
2. Component addition for Cy5 labeling addition for Cy3 labeling.
3. Total RNA (>7 mg/ml) 150–200 μ g 50–80 μ g.
4. Anchored primer (2 μ g/ μ l) 1 μ l 1 μ l
5. DEPC H₂O to 17 μ l to 17 μ l
6. If using an oligo dT(12–18) primer, anneal the primer to the RNA in the following 17 μ l reaction.
7. Component addition for Cy5 labeling addition for Cy3 labeling.
8. Total RNA (>7 mg/ml) 150–200 μ g 50–80 μ g.
9. dT(12–18) primer (1 μ g/ μ l) 1 μ l 1 μ l.
10. DEPC H₂O to 17 μ l to 17 μ l.
11. The incorporation rate for Cy5-dUTP is less than that of Cy3-dUTP, so more RNA is labeled to achieve more equivalent signal from each species.
12. Heat to 65 °C for 10 min and cool on ice for 2 min.
13. Add 23 μ l of reaction mixture containing either Cy5-dUTP or Cy3-dUTP nucleotides, mix well by pipetting, and use a brief centrifuge spin to concentrate in the bottom of the tube:

Reaction mixture (23 μ l)

5 \times first strand buffer	8 μ l
10 \times low T dNTPs mix	4 μ l
Cy5 or Cy3 dUTP (1 mM)	4 μ l
M DTT	4 μ l
Rnasin (30 u/ μ l)	1 μ l
Superscript II (200 u/ μ l)	2 μ l

Superscript polymerase is very sensitive to denaturation at air/liquid interfaces, so be very careful to suppress foaming in all handling of this reaction.

1. Incubate at 42 °C for 30 min. Then add 2 µl Superscript II. Make sure the enzyme is well mixed in the reaction volume and incubate at 42 °C for 30–60 min.
2. Add 5 µl of 0.5 M EDTA.
3. Add 10 µl 1 N NaOH and incubate at 65 °C for 60 min to hydrolyze residual RNA. Cool to room temperature.
4. Neutralize by adding 25 µl of 1 M Tris-HCl (pH 7.5).
5. Desalt the labeled cDNA by adding the neutralized reaction, 400 µl of TE pH 7.5, and 20 µg of human C0t-1 DNA to a MicroCon 100 cartridge. Pipette to mix and spin for 10 min at 500×g.
6. Wash again by adding 200 µl TE pH 7.5 and concentrating to about 20–30 µl (approximately 8–10 min at 500×g).
7. Recover by inverting the concentrator over a clean collection tube and spinning for 3 min at 500×g.

In some cases, the cy5-labeled cDNA will form a gelatinous blue precipitate that is recovered in the concentrated volume. The presence of this material signals the presence of contaminants. The more extreme the contamination, the greater the fraction of cDNA which will be captured in this gel. Even if heat solubilized, this material tends to produce uniform, nonspecific binding to the DNA targets. When concentrating by centrifugal filtration, the times required to achieve the desired final volume are variable. Overly long spins can remove nearly all the water from the solution being filtered. When fluor-tagged nucleic acids are concentrated onto the filter in this fashion, they are very hard to remove, so it is necessary to approach the desired volume by conservative approximations of the required spin times. If control of volumes proves difficult, the final concentration can be achieved by evaporating liquid in the speed vac. Vacuum evaporation, if not to dryness, does not degrade the performance of the labeled cDNA.

1. Take a 2–3 µl aliquot of the Cy5-labeled cDNA for analysis, leaving 18–28 µl for hybridization.
2. Run this probe on a 2 % agarose gel (6 cm wide × 8.5 cm long, 2 mm wide teeth) in Tris Acetate Electrophoresis Buffer (TAE).
3. Scan the gel on a molecular dynamics storm fluorescence scanner (setting: red fluorescence, 200 µm resolution, 1,000 V on PMT).

17.2.6. Hybridization

This protocol describes the conditions for hybridizing fluor-tagged cDNA representations of the mRNA pools of the samples to the EST PCR products immobilized on the glass microarrays.

17.2.6.1. Materials

Microarray Hybridization Chamber

- 50× Denhardt's blocking solution
- Pd (A)40–60 resuspend at 8 mg/ml, and store frozen –20°C (#27-7988, Amersham Pharmacia Biotech.)
- 20× SSC
- Yeast tRNA
- 10 % SDS
- Coverslips
- Forceps
- Coplin jars
- 0.2 ml thinwall PCR tubes
- 65 °C water bath
- Thermocycler for 0.2 ml thinwall PCR tubes
- Microarray scanner
- Image analysis software

Reagents and Solutions

- 0.5× SSC/0.01 % SDS washing buffer.
- Add 25 ml 20× SSC to 974 ml DEPC H₂O.
- Sterile filter on a 0.5 µm filter device.
- Add 1 ml 10 % SDS and mix well.
- Store at room temperature.

0.06× SSC Washing Buffer

- Add 3 ml 20× SSC to 997 ml DEPC H₂O.
- Sterile filter on a 0.5 µm filter device.
- Store at room temperature.

10 mg/ml Human C0t-1 DNA

- Add 925 µl 100 % ethanol and 75 µl 3 M sodium acetate (pH 5.2) to 500 µl Human C0t-1.
- DNA (1 µg/µl).
- Centrifuge at 14,000×g to pellet.
- Aspirate off supernatant and allow to air dry for 5 min.
- Resuspend the pellet in 50 µl DEPC H₂O.

Yeast tRNA

- Resuspend yeast tRNA at 10 mg/ml in DEPC (based on the Supplier's quantitation) in a 1.5 ml polypropylene conical centrifuge tube.
- Add one-half volume of neutralized phenol and vortex.
- Add one-half volume of chloroform and vortex
- Centrifuge 5 min at 10,000×g.
- Transfer aqueous layer to a new 1.5 ml polypropylene conical centrifuge tube.
- Add 1 volume of chloroform and vortex.
- Centrifuge 5 min at 10,000×g.
- Repeat chloroform extraction.
- Transfer aqueous layer to a new 1.5 ml polypropylene conical centrifuge tube.
- Add 0.1 volume of 3 M sodium acetate (pH 5.2).
- Add 2 volumes of ethanol.
- Centrifuge 5 min at 10,000×g.
- Aspirate off supernatant.
- Add 1 volume of 70 % ethanol.
- Centrifuge 5 min at 10,000×g.
- Aspirate off supernatant.
- Allow pellet to dry.
- Resuspend in DEPC water at the original volume.
- Determine the RNA concentration by spectrometry.
- Dilute to 4 mg/ml and store frozen at -20°C .

*17.2.6.2. Steps for
Hybridization and Washing*

1. Determine the volume of hybridization solution required. The rule of thumb is to use $0.033\ \mu\text{l}$ for each mm^2 of slide surface area covered by the coverslip used to cover the array. An array covered by a 24 mm by 50 mm coverslip will require $40\ \mu\text{l}$ of hybridization solution. The volume of the hybridization solution is critical. When too little solution is used, it is difficult to seat the coverslip without introducing air bubbles over some portion of the arrayed ESTs, and the coverslip will not sit at a uniform distance from the slide. If the coverslip is bowed toward the slide in the center, there will be less labeled cDNA in that area and hybridization will be nonuniform. When too much volume is applied, the coverslip will move easily during handling, leading to misplacement relative to the arrayed ESTs, and non-hybridization in some areas of the array.
2. For $40\ \mu\text{l}$ hybridization, pool the Cy3- and Cy5-labeled cDNAs into a single 0.2 ml thin wall PCR tube and adjust

the volume to 30 μl by either adding DEPC H_2O or removing water in a Speed Vac. If using a vacuum device to remove water, do not use high heat or heat lamps to accelerate evaporation. The fluorescent dyes could be degraded.

3. For a 40 μl hybridization combine the following components:

High sample blocking high array blocking

Cy5 + Cy3 probe	30 μl + 28 μl
Poly d(A) (8 mg/ml)	1 μl + 2 μl
Yeast tRNA (4 mg/ml)	1 μl + 2 μl
Human C0t-1 DNA (10 mg/ml)	1 μl + 0 μl
20 \times SSC	6 μl + 6 μl
50 \times Denhardt's blocking solution	1 μl + 2 μl
Total volume	40 μl + 40 μl

4. Mix the components well by pipetting, heat at 98 $^{\circ}\text{C}$ for 2 min in a PCR cycler, cool quickly to 25 $^{\circ}\text{C}$, and add 0.6 μl of 10 % SDS.
5. Centrifuge for 5 min at 14,000 \times g.
6. Apply the labeled cDNA to a 24 mm \times 50 mm glass coverslip and then touch with the inverted microarray.
7. Place the slide in a microarray hybridization chamber, add 5 μl of 3 \times SSC in the reservoir, if the chamber provides one, or at the scribed end of the slide, and seal the chamber. Submerge the chamber in a 65 $^{\circ}\text{C}$ water bath and allow the slide to hybridize for 16–20 h.
8. Remove the hybridization chamber from the water bath, cool, and carefully dry off. Unseal the chamber and remove the slide.
9. Place the slide, with the coverslip still affixed, into a Coplin jar filled with 0.5 \times SSC/0.01 % SDS wash buffer. Allow the coverslip to fall from the slide and then remove the coverslip from the jar with a forceps. Allow the slide to wash for 2–5 min.
10. Transfer the slide to a fresh Coplin jar filled with 0.06 \times SSC. Allow the slide to wash for 2–5 min.
11. Transfer the slide to a slide rack and centrifuge at low rpm (700–1,000) for 3 min in a clinical centrifuge equipped with a horizontal rotor for microtiter plates.

17.2.7. Laser Scanning of a Microarray

This protocol describes how to scan a hybridized slide using a Scanarray 3000 scanner. Turn on the scanner and then start the computer. Open the Scanarray software. Two image windows will

appear inside the Scanarray window, one for each channel. Insertion of slide for reading and click on the eject slide button. Insert the microarray slide array surface up into the slide carriage that appears specification of array extent. Select Start from the Acquire menu; the Acquire Image window will appear. Set the scan size by selecting the Custom radio button. Enter 1.0 mm as the X start position and 17.5 mm as the Y start position to scan. In the X box, enter 20 mm (the width of the array); in the Y box, enter 40 mm (length of array).

Do the following for each channel. Complete both steps for channel 2 (the Cy5 channel) and then repeat with channel 1 (Cy3). Always start with channel 2 in scanning. To assess if laser power and PMT settings for the channel will yield overall intensities covering the range of intensity values (0–65,535), perform a Quick Scan (50 μ m resolution). Set laser power to 65 % of maximum and PMT voltage to 80 % of maximum as an initial approximation. Check the Quick Scan box, select the channel being scanned, and click Acquire. Be sure that only a single channel box is checked; otherwise both channels will be scanned. The image from the channel will appear in one of the image windows as the scan is made.

17.3 Spot Quantization

Spot quantization is the assigning of numerical values to spots imaged by the scanner system. The fluorescence signals in each spot are encoded in a pixellated image file. The spot quantization process is based on standard image processing and recognition technology which has been adapted to find circular spots in a regular grid pattern.

The process begins with the output of a laser scanner system, which is two image files, in tagged image format (TIFF), Windows Bitmap (BMP), or other common image format. The image consists of a grid of pixels, each of which has a 16 bit grayscale (see the Laser Scanning summary). Quantitation begins with the reading of these pixellated images into a quantization software package. Packages currently in use include Biodiscovery's Imagene 3.0 [1], Michael Eisen's freeware package Scanalyze [2], and Imaging Research's ArrayVision [3].

The basic unit of quantization is the microarray spot, typically around 100 μ m in diameter. Scanner resolution is typically 10 μ m, so there are approximately 75 pixels per spot. A well-captured spot should have sharp edges and only a small amount of variation in its individual pixel values. Often, a filtering step is performed by the quantization software to smooth outlier pixel

values (a single intense pixel in a background of low intense pixels, for example); by applying a moving median or average filter. The quality of the pixellated spot on the image ultimately depends on the physical spotting process, how well the spot DNA was cross-linked to the slide substrate, and/or how well mixed the hybridization solution was upon application to the spot DNA. In order to compare Cy3 and Cy5 signal in a given spot, the pixels in its Cy3 image must be matched with the corresponding pixels in the Cy5 image. The software package will have a feature to align each spot's Cy3 and Cy5 image. Features will be provided to translate, rotate, shrink, and expand one image relative to the other, to obtain accurate superposition.

17.4 Data Normalization and Statistical Analysis

Before it is possible to draw biological conclusions or to apply sophisticated statistics, it is important to normalize the data. This corrects for systematic biases resulting basically from different amounts of RNA used for labeling, different incorporation efficiencies of the Cy3 and Cy5 dyes in the labeling protocols, and different detection efficiencies of the dyes [9, 10]. The data obtained are normalized using the software on the assumption that the log transformed data approach normal distribution. Usually the global normalization method and the Cy5 and Cy3 intensities are related by a constant factor ($Cy5 = kCy3$) and the expression ratio of the average or median gene of the population is zero ($Cy5/Cy3$ ratio for an average gene is one). The average gene is based on the observation that the majority of the genes do not change their expression levels when comparing two mRNA populations. Significance analysis of microarrays (SAM) is a statistical technique to determine whether changes in gene expression are statistically significant. With the advent of DNA microarrays it is now possible to measure the expression of thousands of genes in a single hybridization experiment. The data generated are considerable and a method for sorting out what is significant and what is not is essential. SAM is distributed by Stanford University in an R-package. SAM identifies statistically significant genes by carrying out gene-specific t-tests and computes a statistic d_j for each gene j , which measures the strength of the relationship between gene expression and a response variable [11, 12]. This analysis uses nonparametric statistics, since the data may not follow a normal distribution. The response variable describes and groups the data based on experimental conditions. In this method, repeated permutations of the data are used to determine if the expression of any gene is significant related to the response. The use of

permutation-based analysis accounts for correlations in genes and avoids parametric assumptions about the distribution of individual genes. This is an advantage over other techniques (for example, ANOVA and Bonferroni), which assume equal variance and/or independence of genes. SAM is available for download online at <http://www-stat.stanford.edu/~tibs/SAM/> for academic and non-academic users.

17.5 Applications of Microarray Technology

17.5.1. Gene Discovery

DNA Microarray technology helps in the identification of new genes and knowing about their functioning and expression levels under different conditions.

17.5.2. Disease Diagnosis

DNA Microarray technology helps researchers learn more about different diseases such as heart diseases, mental illness, infectious disease, and especially the study of cancer. Until recently, different types of cancer have been classified on the basis of the organs in which the tumors develop. Now, with the evolution of microarray technology, it will be possible for the researchers to further classify the types of cancer on the basis of the patterns of gene activity in the tumor cells. This will tremendously help the pharmaceutical community to develop more effective drugs as the treatment strategies will be targeted directly to the specific type of cancer.

17.5.3. Drug Discovery

Microarray technology has extensive application in Pharmacogenomics [13, 14]. Pharmacogenomics is the study of correlations between therapeutic responses to drugs and the genetic profiles of the patients. Comparative analysis of the genes from a diseased and a normal cell will help the identification of the biochemical constitution of the proteins synthesized by the diseased genes. The researchers can use this information to synthesize drugs which combat with these proteins and reduce their effect.

17.5.4. Toxicological Research

Microarray technology provides a robust platform for the research of the impact of toxins on the cells and their passing on to the progeny. Toxicogenomics establishes correlation between responses to toxicants and the changes in the genetic profiles of the cells exposed to such toxicant.

17.5.5. Microarray in Microbiology

Microarrays can be used in microbiology for a multitude of differing applications, from the study of gene regulation and bacterial response to environmental changes, genome organization, and

evolutionary questions up to taxonomic and environmental studies. The knowledge of the main aspects of this technology helps to understand these specific applications. Vital for further advances of microarray technology in microbiology will be the recognition of the importance of the physiological experiments ahead of the transcription analysis, the standardization of protocols and controls for transcription analysis, more integration of the data analysis with biochemical and genetic knowledge, and flexible and intuitive databases for mining the vast amounts of data [15].

References

1. Wildsmith SE, Elcock FJ (2001) Microarrays under the microscope. *Mol Pathol* 54:8–16
2. DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 14:457–460
3. Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470
4. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM (1999) Expression profiling using cDNA microarrays. *Nat Genet* 21:10–14
5. Heller RA, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J, Woolley DE, Davis RW (1997) Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc Natl Acad Sci USA* 94:2150–2155
6. Affymetrix GeneChip technology, <http://www.affymetrix.com/technology/index.affx>. An overview of how an Affimatrix GeneChip is made, <http://www.affymetrix.com/technology/manufacturing/index.affx>.
7. Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephanians SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 19:342–347
8. Lander ES (1999) Array of hope. *Nat Genet* 21:3–4
9. Yang IV, Chen E, Hasseman JP, Liang W, Frank BC, Wang S, Sharov V, Saeed AI, White J, Li J, Lee NH, Yeatman TJ, Quackenbush J (2002) Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome Biol* 3:0062
10. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30:e15
11. Kulesh DA, Clive DR, Zarlenga DS, Greene JJ (1987) Identification of interferon-modulated proliferation-related cDNA sequences. *Proc Natl Acad Sci USA* 84:8453–8457
12. Lausted C, Dahl T, Warren C, King K, Smith K, Johnson M, Saleem R, Aitchison J, Hood L, Lasky SR (2004) POSaM: a fast, flexible, open-source, inkjet oligonucleotide synthesizer and microarrayer. *Genome Biol* 5:R58. doi:10.1186/gb-2004-5-8-r58. PMID 15287980
13. Jain K (2000) Applications of biochip and microarray systems in pharmacogenomics. *Pharmacogenomics* 1:289–307
14. Crowther DJ (2002) Applications of microarrays in pharmaceutical industry. *Curr Opin Pharmacol* 2:551–554
15. Ehrenreich A (2006) DNA microarray technology for the microbiologist: an overview. *Appl Microbiol Biotechnol* 73:255–273

Chapter 18

Microarray Analysis of Different Functional Genes of Microorganisms

Hirak Ranjan Dash and Surajit Das

Abstract

Microorganisms play an integral and unique role in ecosystem function and sustainability. Understanding the structure and composition of microbial communities and their responses and adaptations to environmental perturbations such as toxic contaminants, climate changes, and agricultural and industrial practices is critical in maintaining or restoring desirable ecosystem functions. The DNA microarray (or microchip) technology is a powerful tool for studying gene expression and regulation on a genomic scale and detecting genetic polymorphism in both prokaryotes and eukaryotes. Compared to conventional membrane-based hybridization, glass slide-based microarrays offer the additional advantages of rapid detection, lower cost, automation, and low background levels. Microarray-based genomic technique is potentially an extremely powerful tool for characterizing microbial communities and their biological functions. Hence, in this chapter, we will focus on the detection of expression level of certain functional genes in *Pseudomonas* spp., *Staphylococcus* spp., and viruses.

18.1 Introduction

An array is an orderly arrangement of samples where matching of known and unknown DNA samples is done based on pairing rules. An array experiment makes use of common assay systems such as microplates or standard blotting membranes. The sample spot sizes are typically less than 200 μm in diameter and usually contain thousands of spots. Thousands of spotted samples known as probes (with known identity) are immobilized on a solid support (a microscope glass slides or silicon chips or nylon membrane). The spots can be DNA, cDNA, or oligonucleotides. These are used to determine complementary binding of the unknown sequences thus allowing parallel analysis for gene expression and gene discovery. An experiment with a single DNA chip can provide information on thousands of genes simultaneously. An orderly arrangement of the probes on the support is important as the location of each spot on the array is used for the identification of a gene. DNA microarray can also be applied in research for

gene expression profiling, that is, identification of changes in mRNA expression of strains exposed to particular substrate, for example, a specific xenobiotics. In microarray technique sequence information is required to design probes. However this approach cannot be applicable for discovering new catabolic genes for which no sequences are available in the databases. Moreover, knowledge of the entire sequence is not necessary for the construction of microarrays, and PCR products of a random genomic library constructed from a microorganism of interest may be used. It is expected that to a toxic substrate, differential gene expression would result at the transcript level. This is reflected by differential hybridization patterns in the presence or absence of the toxic pollutant. Afterward, clones of the library associated with differentially hybridized probes can be picked up for sequencing. To analyze the expression level of various functional genes in different microorganisms, the basic protocol is the same in different organisms; however, the gene chip where the oligonucleotides are immobilized varies, which is specific for that specific organism. Microarray can be applied to analyze the xenobiotic degrading genes in environmental bacteria, toxic genes in clinical bacteria, and expression level of various genes in host due to the infection of a specific virus.

Pseudomonas are gram-negative, motile, aerobic, non-spore-forming bacteria demonstrating great deal of metabolic diversity possessing the characteristic features of opportunistic pathogens to human beings, biofilm formation, and are responsible in the process of bioremediation. *Pseudomonas* species have been reported to be used in the process of bioremediation of polycyclic aromatic hydrocarbons, toluene, carbazole, aromatic organic compounds, carbon tetrachloride, and almost all heavy metals [1]. As *Pseudomonas* possess multifaceted properties in bioremediation point of view, it can be used as a potential model organism for the study of microbial bioremediation.

Staphylococcus species are potential cause of various diseases in human causing skin infection, eye infection, and abdominal infection which can lead to organ damage, damage to the eye, or even death. Though some of the strains of *Staphylococcus* are the normal microflora of the body, they can change their role by becoming opportunistic pathogens. *Staphylococcus* are of upmost importance due to their rapid acquisition of antibiotic resistance properties as Methicillin resistant *Staphylococcus aureus* (MRSA) or Vancomycin resistant *Staphylococcus aureus* (VRSA) [2]. Microarray analysis of *Staphylococcus* can reveal the diagnosis of these organisms at genus and species level and the molecular mechanism antibiotic resistance by detecting various genes.

Microarray has been evolved as a potential tool for the diagnosis of a viral infection and to study the expression level of various genes in the host due to a particular viral infection [3]. Due to

infection of a specific virus, the regulation patterns of genes are changed in the host. Targeting the detection of genes that are upregulated or downregulated in host, the cause of viral infection and the diagnosis of the disease becomes easier. In some cases, for the identification of a particular virus, the gene chips are constructed using the oligonucleotides of the genome of the virus and by targeting the genes of the specific virus. Analysis of viruses like HIV and influenza virus using Microarray is common nowadays.

18.2 Materials

18.2.1. Total RNA Isolation

1. QIAGEN^R RNeasy Mini Purification Kit

18.2.2. Preparation of Poly-A Controls

1. Poly-A RNA control kit

18.2.3. cDNA Synthesis

1. RNA/primer hybridization mix
2. 5× 1st standard buffer
3. 100 mM DTT
4. 10 mM dNTPs
5. SUPERase In (20 U/μl)
6. SuperScript II (200 U/μl)

18.2.4. Removal of DNA

1. 1 N NaOH
2. 1 N HCl

18.2.5. Purification and Quantification of cDNA

1. MiniElute PCR purification columns
2. Spectrophotometer

18.2.6. cDNA Fragmentation

1. 10× DNase I Buffer
2. cDNA
3. DNase I
4. Nuclease free water

18.2.7. Terminal Labeling

1. 5× reaction buffer
2. GeneChip DNA labeling reagent, 7.5 mM
3. Terminal deoxynucleotidyl transferase
4. Fragmentation cDNA product
5. Water

18.2.8. Target hybridization, Washing, Staining, and Scanning

1. GeneChip hybridization, wash, and stain kit
2. Hybridization module, Box1
3. 2X Hybridization mix
4. DMSO
5. Nuclease free water
6. Control Oligo B2, 3 nM
7. Hybridization Oven 640
8. Sterile, RNase free, microcentrifuge vials, 1.5 ml
9. Micropipette
10. Sterile barrier pipette tips and non-barrier tips
11. Stain Cocktail 1
12. Stain Cocktail 2
13. Array holding buffer
14. Wash buffer A and B
15. RNase free water
16. Fluidics Station 450
17. GeneChip[®] Scanner 3000
18. Tough-Spots[™], Label Dots
19. Experiment and Fluidics Station Setup

18.3 Method

18.3.1. Total RNA Isolation

1. Bacterial culture were grown in liquid basal media and the overnight culture was harvested by centrifugation.
2. Total RNA can be isolated by using standard procedures for bacterial RNA isolation QIAGEN[®] RNeasy Mini Purification Kit.

18.3.2. Preparation of Poly-A RNA Controls

1. The Poly-A RNA Control Stock and Poly-A Control Dilution buffer are generally provided with the Poly-A RNA Control Kit to prepare the appropriate serial dilutions based on the following recommendations (Table 18.1).

18.3.3. cDNA Synthesis

1. Prepare the following mixture for primer annealing.
2. Incubate the RNA/Primer mix at the following temperatures.
 - 70 °C for 10 min
 - 25 °C for 10 min
 - Chill to 4 °C

Table 18.1
Serial dilutions of poly-A RNA control stock

Array format	Serial dilution		Spike-in volume (μl)
	First	Second	
169 Format (Mini)	1:20	1:16	2
100 Format (Midi)	1:20	1:20	2
49 Format (Standard)	1:20	1:13	2

Table 18.2
Ingredients for cDNA synthesis

Ingredients	Volume (μl)	Final dilution
RNA/Primer hybridization mix (from previous step)	30	–
5 × 1st Standard buffer	12	1 ×
100 mM DTT	6	10 mM
10 mM dNTPs	3	0.5 mM
SUPERase In (20 U/μl)	1.5	0.5 U/μl
SuperScript II (200 U/μl)	7.5	25 U/μl
Total volume	60	

3. Prepare the reaction mix for cDNA synthesis. Briefly centrifuge the reaction tube to collect sample at the bottom and add the cDNA synthesis mix from Table 18.2 to the RNA/primer hybridization mix
4. Incubate the reaction at the following temperatures:
 - 25 °C for 10 min
 - 37 °C for 60 min
 - 42 °C for 60 min
 - Inactivate SuperScript II at 70 °C for 10 min
 - Chill to 4 °C

18.3.4. Removal of RNA

1. Add 20 μl of 1 N NaOH and incubate at 65 °C for 30 min.
2. Add 20 μl of 1 N HCl to neutralize.

Table 18.3
Fragmentation reaction

Ingredients	Volume	Dilution
10× DNase I Buffer	2 µl	1×
cDNA	10 µl	–
DNase I	X µl	0.6 U/µg of cDNA
Nuclease free water	Up to 20 µl	–
Total volume	20 µl	

**18.3.5. Purification
and Quantification
OF cDNA**

1. Quantify the purified cDNA product by 260 nm absorbance.
2. Use MiniElute PCR Purification Columns to clean up the cDNA synthesis product. Elute the product with 12 µl of EB. The average volume of the elute is 11 µl from 12 µl of elution buffer.
3. Typical yield of cDNA are 3–7 ng. A minimum of 1.5 µg of cDNA is required for subsequent procedures to obtain sufficient material to hybridize onto the array and to perform necessary quality control experiments.

**18.3.6. cDNA
Fragmentation**

1. Prepare the following reaction mix (Table 18.3).
2. Incubate the reaction at 37 °C for 10 min.
3. Inactivate DNase I at 98 °C for 10 min.
4. The fragmented cDNA is applied directly to the thermal labeling reaction.
5. Alternatively, the material can be stored at –20 °C for later use.

**18.3.7. Terminal
Labeling**

1. Prepare the following reaction mix (Table 18.4):
2. Incubate the reaction at 37 °C for 60 min.
3. Stop the reaction by adding 2 µl of 0.5 M EDTA.
4. The target is ready to be hybridized onto probe arrays. Alternatively it may be stored at –20 °C for later use.

**18.3.8. Target
Hybridization,
Washing, Staining
and Scanning**

1. Prepare the following hybridization cocktail.
2. Equilibrate probe array to room temperature immediately before use.

Table 18.4
Ingredients for terminal label reaction

Ingredients	Volume
5 × Reaction buffer	10 µl
GeneChip DNA labeling reagent, 7.5 mM	2 µl
Terminal deoxynucleotidyl transferase	2 µl
Fragmentation cDNA product	Up to 20 µl
Water	16 µl
Total volume	50 µl

- Based on the format of the array type used, add the appropriate volume of hybridization cocktail.

Array	Volume
49 Format (Standard)	200 µl
100 Format (Midi)	130 µl
169 Format (Mini)	80 µl

- Place probe array in the hybridization oven set at 50 °C.
- To avoid stress to the motor, load probe arrays in a balanced configuration around axis. Rotate at 60 rpm.
- Hybridize for 16 h. Prepare reagents for the washing and staining steps required immediately after completion of hybridization.

Step 1: Defining File Locations

- Launch Microarray Suite from the workstation and select. Tools → Defaults → File Locations from the menu bar
- The file Locations windows displays the locations of the following files:
 - Probe Information (library files, mask files)
 - Fluidics Protocols (fluidics station scripts)
 - Experiment Data (.exp, .dat, .cel, and .chp fil
- Verify that all three file locations are set correctly and click OK.

Step 2: Entering Experiment Information

To wash, stain and scan a probe array, an experiment must first be registered in GCOS or Microarray Suite. The field of information required for registering experiments in Microarray Suite are:

- Experiment Name
- Probe Array type
- Sample Name
- Sample Type
- Project

Step 3: Preparing the Fluidics Station

The Fluidics Station 400 or 450/250 is used to wash and stain the probe arrays. It is operated using GCOS/Microarray Suite.

1. Turn on the fluidics Station using the switch on the lower left side of the machine.
2. Select Run → Fluidics from the menu bar.
3. Priming should be done after that:
 - When the fluidics station is first started
 - When wash solutions are changed
 - Before washing if a shutdown has been performed
 - If the LCD window instructs the user to prime

Preparing the stain reagents

1. Remove Stain Cocktail 1, Stain Cocktail 2, and array Holding buffer from the stain module.
2. Gently tap the bottles to mix.
3. Aliquot the following reagents:
 - (a) 600 μ l of Stain Cocktail 1 into a 1.5 ml amber microcentrifuge vial.
 - (b) 600 μ l of Stain Cocktail 2 into a 1.5 ml clear microcentrifuge vial.
 - (c) 800 μ l of Array Holding Buffer into a 1.5 ml clear microcentrifuge vial.
4. Spin down all vials to remove the presence of any air bubbles.

Washing and Staining the Probe Array

1. In the fluidics Station dialog box on the work station, select the correct experiment name from the drop-down experiment list. The probe Array type appears automatically.

2. In the protocol drop-down list, select the appropriate anti-body amplification protocol to control the washing and staining of the probe array format being used.
3. Choose Run in the Fluidics Station dialog box to begin the washing and staining. Follow the instructions in the LCD window.
4. Insert the appropriate probe array into the designated module of the fluidics station while the cartridge lever is down or in the eject position. When finished, verify that the cartridge lever is returned to the up or engaged position.
5. Remove any microcentrifuge vials remaining in the sample holder of the fluidics station module being used.
6. Follow the instructions on the LCD window of the fluidics station by placing the three experiment sample vials into the sample holders 1, 2, and 3 on the fluidics station.
7. When the protocol is complete, the LCD window displays the message EJECT & INSPECT CATRIDGE.
8. Remove the probe arrays from the fluidics station modules by first pressing down the cartridge lever to the eject position.
9. Check the probe array window for large bubbles or air pockets. If the probe array has no large air bubbles, it is ready to scan on the GeneArray[®] Scanner or the GeneChp[®] Scanner 3000.
10. Keep the probe arrays at 4 °C and in dark until ready for scanning.
11. If there are no more samples to hybridize, shut down the fluidics station.

Scanning the Probe Array

1. Select Run Scanner from the menu bar. Alternatively, click the start Scan icon in the tool bar. The scanner dialog box appears with a drop-down list of experiments that have not been run.
2. Select the experiment name that corresponds to the probe array to be scanned. A previously run experiment can also be selected by using the Include Scanned Experiments option box. After selecting this option, previously scanned experiments appear in the drop-down list.
3. Once the experiment has been selected, click the start button. A dialog box prompts you to load the sample into the scanner.
4. Open the sample door on the scanner and insert the probe array into the holder. Do not force the probe array into the holder. Close the sample door of the scanner. If you are using

the GeneChip Scanner 3000, do not attempt to close the door by hand.

5. Click OK in the Start Scanner dialog box. The scanner begins scanning the probe array and acquiring data. When Scan in progress is selected from the View menu, the probe array image appears on the screen as the scan progresses.

References

1. Ojo OA (2007) Molecular strategies of microbial adaptation to xenobiotics in natural environment. *Biotechnol Mol Bio Rev* 2:001–013
2. Volokhov D, Chizhikov V, Chumakov K, Rasooly A (2003) Microarray analysis of erythromycin resistance determinants. *J Appl Microbiol* 95:787–798
3. Raaben M, Whitley P, Bouwmeester D, Setterquist RA, Rottier PJM, Hann CAM (2008) Improved microarray gene expression profiling of virus-infected cells after removal of viral RNA. *BMC Genomics* 9:12

Chapter 19

DNA Cloning and Sequencing

Bandamaravuri Kishore Babu, Anu Sharma, and Hari Kishan Sudini

Abstract

The plasmid DNA is cleaved with an enzyme and joined in vitro to foreign DNA; the resulting recombinant plasmids are then used to transform bacteria. The plasmid vectors must be carefully chosen and processed to minimize the effort required to identify and characterize recombinants. This chapter provides guidelines for preparation of DNA fragment for cloning, transformation into chemically competent host, and selection of positive clones. The write-up will also describe basic methods used in the cloning of PCR amplified rRNA gene into appropriate vector and followed by sequencing.

19.1 Introduction

In principle, cloning in plasmid vectors is very straightforward. The easiest fragment to clone carries a noncomplementary protruding termini generated by digestion with two different restriction enzymes. Since most of the present-day vectors contain poly-linker that has multiple cloning sites, it is almost always possible to find restriction sites that are compatible with the termini of the foreign DNA fragment. The fragment of foreign DNA is then inserted into the vector by a process known as directional cloning.

Fragment of foreign DNA carrying identical termini (either blunt-ended or protruding) must be cloned in a linearized plasmid vector bearing compatible ends. During the ligation reaction, the foreign DNA and the plasmid DNA should have the capacity to circularize and to form tandem oligomers. It is therefore necessary to carefully adjust the concentrations of the two types of DNA in the ligation reaction to optimize the number of correct ligation products. In addition, removal of the 5'-phosphate groups with alkaline phosphatase will help to suppress self-ligation and circularization of the plasmid DNA.

19.2 Materials

1. Glassware, instruments, and other materials: Electrophoresis tank (horizontal/submarine), UV transilluminator, sterile microfuge tubes, micropipettes, sterile micropipette tips, shaker incubator, sterile 50 ml polypropylene centrifuge, ice-box, microfuge, vortex mixture, water bath, etc.
2. Chemicals:
 - (a) Agarose
 - (b) 50× TAE buffer
 - (c) Ethidium bromide (10 mg/ml)
 - (d) Restriction endonucleases (*Eco*RI, *Hind*III, etc.)
 - (e) λ DNA digested with *Hind*III
 - (f) Isolated plasmids (with and without inserts of foreign DNA)
 - (g) Calf Alkaline Phosphatase (CIP)
 - (h) T4 DNA ligase and 10× ligase buffer
 - (i) Buffer saturated phenol
 - (j) Chloroform
 - (k) 3 M Sodium acetate (pH 7)
 - (l) 3 M Sodium acetate (pH 5.2)
 - (m) 10 M Ammonium acetate
 - (n) LB broth
 - (o) LB agar plates with and without ampicillin
 - (p) *E. coli* DH5 α Competent cells
 - (q) IPTG (200 mg/ml)
 - (r) X-gal (20 mg/ml in dimethylformamide)
 - (s) Ptz57R/T vector (MBI Fermentas) or PCRTM II vector (Invitrogen)
 - (t) Calcium chloride—0.2 M (prepare 1 M stock solution of CaCl₂ and store 10 ml aliquots frozen at -20°C). Just before use, dilute an aliquot to 100 ml with sterile water and by filter through 0.45 μm filter, and then chill on ice for use.
 - (u) BDT v3.1 Reaction Mix (Applied Biosystems #4337455) [1]
 - (v) 5× Sequencing buffer (Applied Biosystems #4336697)
 - (w) Hi-Di formamide (Applied Biosystems)
 - (x) 0.5 M EDTA pH 8.0
 - (y) 125 mM EDTA
 - (z) TE buffer

19.3 Methods

19.3.1. Ligation of Vector and Foreign DNA Fragment

Ligation of a segment of foreign DNA to a linearized plasmid vector involves the formation of new bonds between phosphate residues located at the 5'-hydroxyl moieties. Ligation of one end of DNA to another can be regarded as a bimolecular reaction whose velocity under standard conditions is determined solely by the concentration of same DNA molecule (intramolecular ligation) or on different molecules (intermolecular ligation). Low concentration of DNA in the ligation reaction may lead to intramolecular ligation, whereas high DNA concentration may result in the formation of dimers and/or larger oligomers of the plasmid.

19.3.2. Preparation of Vector DNA and Foreign DNA Fragment for Cloning

1. Restriction digestion of plasmid DNA (prepared by mini prep.) and foreign DNA with the desired endonucleases.
2. Check 5 µl of the above plasmid for completion of digestion in 0.8 % agarose gel. Use undigested plasmid and marker DNA for comparison.
3. Mix following components in the order mentioned below in a microfuge tube on ice:

To linearize the plasmid:

Plasmid DNA (1–2 µg)	5 µl
10× RE buffer	2 µl
Restriction enzyme 1 (e.g., <i>Eco</i> RI)	2 µl
Restriction enzyme 2 (e.g., <i>Hind</i> III)	2 µl
Deionized distilled water	to 20 µl

4. Mix the contents by gentle tapping and pulse spin in a microfuge to bring down all the liquid to the bottom of the tube and incubate at 37 °C for 1 h.
5. Heat at 65 °C for 10 min to stop the reaction. Chill on ice. Add 5 µl of loading dye, mix, and pulse spin. Load the digested products on 1 % TAE/agarose gel, with λ DNA digested with *Hind* III as marker.

19.3.3. Preparation of Phosphatase-Treated Vector

1. Restriction digested plasmid will be now proceeding for phosphatase treatment as follows:

Digested plasmid	15 μ l
10 \times CIP buffer	5 μ l
H ₂ O	49 μ l
Calf alkaline phosphatase (1 U)	1 μ l

2. Incubate at 37 °C for 30 min.
3. Add another 1 μ l of CIP and continue the incubation for 30 min.
4. After 1 h of CIP treatment, add 1 μ l of 0.5 M EDTA (pH 8.0) to get a final concentration of 5 mM.
5. Incubate at 75 °C for 10 min to inactivate the CIP.
6. Cool the reaction to room temperature and extract once with equal volume of phenol and once with equal volume of phenol–chloroform mixture.
7. To the aqueous phase, add 0.1 volume of 3 M sodium acetate (pH 7.0), and 2 volume of ethanol, mix well, and precipitate the linear dephosphorylated vector at –20 °C for 30 min.
8. Recover the DNA by centrifugation at 4 °C in a microfuge.
9. Wash the pellet with 70 % ethanol, air dry, and dissolve in 10 μ l of distilled water.
10. Run an aliquot of both the vector and the insert DNA to estimate the concentration of DNA in the gel before setting up the ligation.

19.3.4. Ligation

1. Set up ligation reaction in total volume of 10 μ l as follows (for a foreign DNA fragment that has length equal to vector DNA) (Table 19.1).
2. If the foreign DNA is smaller than vector, reduce the concentration of foreign DNA accordingly to bring it equal to molar concentration of the vector.
3. Include necessary control like ligation without foreign DNA fragment (Vector re-circularization control), ligation vector only that was not treated with phosphatase (Ligation control), etc.
4. In each case, adjust the volume to 10 μ l with phosphatase (Ligation control), etc.
5. In each case, adjust the volume to 10 μ l with H₂O.
6. Incubate the reaction for 4–16 h at 16 °C.
7. Use 2 μ l of the ligation mixture for transformation of bacteria. Store the remaining ligation reaction at –20 °C for further use.

Table 19.1
Ligation reaction mixture of 10 l volume

Ingredients	Volume	Final concentration
Vector DNA (100 ng/ μ l)with dephosphorylated 5'termini	2 μ l	200 ng
Foreign DNA fragment (100 ng/ μ l) with compatible phosphorylated termini	2 μ l	200 ng (equimolar to vector)
10 \times ligase buffer with 10 mM ATP	1 μ l	1 \times
T4 DNA ligase	1 μ l	0.1 Weiss Unit
Deionized distilled water	To 10 μ l	0.1 Weiss Unit

19.3.5. Transformation

When the bacteria are treated with ice-cold solution of CaCl_2 and then briefly heated, they could be transfected with plasmid DNA. Apparently the treatment induces a transient state of “competence” in the recipient bacteria, during which they are able to take up DNAs derived from a variety of sources. Most commonly used methods yield transformants at a frequency of 10^7 – 10^8 transformants/ μ g of supercoiled plasmid. Competent cells of the *E. coli* strains such as JM 107, XL 1-Blue, DH5 α , SURE, and NM522 can be used for transformation.

19.3.6. Competent Cell Preparation

Day 1: Selection of *E. coli* DH5 α on LB Agar Plates

1. Streak *E. coli* DH5 α culture either from glycerol stock or from any viable culture stored at 4 °C, with platinum loop on LB agar plates.
2. Incubate the LB plate at 37 °C overnight (O/N) and isolate a single colony.

Day 2: Preparation for O/N Culture

1. Inoculate 5 ml of LB broth with a single colony from LB agar plate.
2. Let it be grown for O/N in a shaking incubator (150 rpm) at 37 °C.
3. Keep the following in the cold room for next day use: 50 ml centrifuge tubes, microfuge tubes, 10 and 5 ml glass pipettes, CaCl_2 solution.

Day 3: Preparation of Competent Cells

1. Inoculate 100 ml LB medium in 1 liter flasks with O/N culture (make 1 % inoculums).

2. Let the flask grow at 37 °C shaking for 2–3 h. Measure OD at A₅₅₀ after every 30 min to find out when A₅₅₀ is between 0.4 and 0.5 (i.e., when the cells are in early log phase).
3. Chill the flasks in ice for 5–10 min. Transfer the culture to pre-chilled 50 ml centrifuge tubes.
4. Centrifuge at 5,000 rpm/5 min at 4 °C.
5. Decant the supernatant and suspend the pellet gently with the help of a pre-chilled glass pipette in 20 ml of 200 mM CaCl₂ and incubate in ice water for 20 min. Centrifuge at 5,000 rpm/5 min at 4 °C.
6. Decant the supernatant and suspend the pellet gently in 4 ml 80 mM CaCl₂ solution.
7. Aliquot the cells in microfuge tubes either in 200 µl or its multiples, and immediately freeze in liquid nitrogen and store the cells at –70 °C.
8. If not used immediately, the competent cells may be stored on ice O/N.

19.3.7. Transformation

1. Thaw a 200 µl aliquot of competent cells on ice.
2. Add 2 µl of ligation mix containing approximately 50–100 ng of DNA to the competent cells.
3. The volume of DNA/ligation mix to be added to a 200 µl aliquot of competent cells should not exceed 10 µl.
4. Mix the contents of tube by swirling gently. Incubate on ice for 30 min.
5. Heat shock at 42 °C for exactly 120 s. (Do not shake the tubes.)
6. Rapidly transfer the tubes to ice bath and allow the cells to chill for 1–2 min.
7. Add four volumes (0.8 ml) of LB medium and keep in a water bath at 37 °C for 1 h to allow the bacteria to recover and express the antibiotic resistance marker encoded by the plasmid.
8. Centrifuge at 5,000 rpm for 5 min to settle down the cells. Remove medium and resuspend the cells in the remaining medium and plate on the LB ampicillin plates with IPTG and X-gal.
9. Include following control:
 - (a) No DNA control
 - (b) Transformation efficiency control
 - (c) Incubate the plates at 37 °C O/N

19.3.8. Selection of Positive Clones

There are four following methods that are commonly used to identify bacterial colonies that contain recombinant plasmids:

19.3.8.1. Restriction Analysis

In restriction analysis, a number of independently transformed bacterial colonies are picked and grown in small cultures. Plasmid DNA isolated from each culture is analyzed by digestion with restriction enzyme and gel electrophoresis.

19.3.8.2. Insertional Inactivation

It can only be used with order vectors (Pbr322) that carry two or more antibiotic resistance genes and an appropriate distribution of restriction enzyme cleavage sites. The foreign DNA is cloned in the plasmid in such a way that it disrupts the reading frame of one of the antibiotic resistant genes. Recombinant bacteria are screened by growing identical colonies separately on more than one antibiotic plate. If the bacteria become sensitive to the antibiotic, whose gene was disrupted by the insertion of the foreign DNA, and remains resistant to the others, it indicates that the bacteria contain the plasmid having foreign DNA.

19.3.8.3. Screening by Colony Hybridization

In this method the bacterial colonies are transferred onto a nitrocellulose paper and lysed to release and denature the DNA that is immobilized on the nitrocellulose paper. This nitrocellulose paper is then hybridized with radioactive labeled DNA fragments that were used for cloning. The autoradiogram is aligned with the original bacterial plate, from where the colonies were transferred, to identify the bacteria carrying foreign DNA.

19.3.8.4. α -Complementation

Many of the vectors in current (e.g., the pUC Series) carry a short segment of *E. coli* DNA that contains the regulatory sequences and the coding information for the first 146 amino acids of the β -galactosidase gene (*lac Z*). Embedded in this coding region is polycloning or multiple cloning site that does not disrupt the reading frame but results in the harmless interpolation of a small number of amino acids to the amino-terminal fragment of β -galactosidase. Vectors of this type are used in host cells that code for the carboxy-terminal portion of the β -galactosidase. Although neither the host-encoded nor the plasmid-encoded fragments are themselves active, they can associate to form enzymatically active protein. This type of complementation, in which deletion mutants of the operator-proximal segment of the *lac-Z* gene are complemented by β -galactosidase—negative mutants that have the operator-proximal region intact, is called α -complementation. The *lac*⁺ bacteria that result from α -complementation are easily recognized because they form blue colonies in the presence of chromogenic substrate 5-bromo-4-chloro-3-indolyl- β -D-galactose (X-gal). However insertion of a fragment of foreign DNA into the polycloning site of the plasmid almost invariably results in the production of an amino-terminal fragment that is not capable of α -complementation.

Bacteria carrying recombinant plasmid therefore form white colonies. The development of this simple color test has greatly simplified the identification of recombinants constructed in plasmid vectors of this type. It is easily possible to screen many thousands of colonies visually and to recognize colonies that carry putative recombinant plasmids. The structure of these plasmids is then verified by restriction analysis of mini preparation of plasmid DNA.

19.3.9. Experimental Procedure

1. To a premade LB agar plate containing ampicillin, add 40 μ l of stock solution of X-gal (20 mg/ml) and 4 μ l of solution of IPTG (200 mg/ml).
2. Using a sterile glass spreader, spread the solution over the entire surface of the plate.
3. Incubate the plate at 37 °C until the fluid has disappeared.
(**Note:** It may take up to 2–3 h, if the plate is freshly made.)
4. Inoculate the plate with 200 μ l of transformation mixture and spread the solution over the entire surface of the entire plate using a sterile glass spreader.
5. Incubate the inoculated plate in an inverted position for 12–16 h at 37 °C.
6. Store the plate at 4 °C for several hours. This allows the blue color to develop fully.

Pick few blue and few white colonies for analysis by restriction endonuclease digestion.

Note: Colonies that contain active β -galactosidase are pale blue in the center and dense blue at the periphery. White colonies occasionally show a faint blue spot in the center, but these are colorless at the periphery.

19.3.10. Cloning of PCR Product Using T/A Over Hang

Cloning of PCR product into appropriate vector followed by sequencing allows the product identification and characterization. The basic methods used in cloning of PCR product include:

TA cloning: Since PCR product generated by *Taq* polymerase is appended with a single extraneous dA at 3' ends, the easiest way of cloning is by using a plasmid tailed with dT.

Blunt end cloning: The blunt end PCR product generated by *Pwo* or *Pfu* polymerase can be cloned into a plasmid restricted with blunt end generating enzymes.

Directional cohesive end cloning: In this case PCR product is first restricted with appropriate restriction enzymes followed by ligating them onto plasmid linearized by same restriction enzymes.

Table 19.2
Components of ligation reaction mixture

Ingredients	Volume (μl)
Plasmid vector pTZ57R/T vector (0.165 μg)	1
PCR fragment (approx 0.495 μg)	2
10× ligation buffer	1
PEG 4000 solution	1
Deionized water	5
T4 DNA Ligase, 5 U	1

19.3.11. Ligation

The following protocol describes cloning of PCR amplified product using TA vector:

1. PCR products are first purified to remove enzymes, unused primers, dNTPs, etc. For this the PCR product is first run on low melting agarose gel, followed by extraction and purification (many commercial gel extraction kits are available).

Note: The efficiency of ligation is known to be dependent on the purity of the PCR fragments and if a single homogenous band of desired size is observed on the gel after purification of PCR product, it can be directly used in the ligation reaction. tpb 6pt

2. Dissolve the purified PCR fragment in 10–20 μl of TE buffer, determine the DNA concentration using nanodrop UV spectrophotometer, or alternatively load 2 μl of DNA into agarose gel electrophoresis and compare with the known amount of DNA markers.
3. Calculate the amount of PCR fragment required for ligation (in equimolar concentration) using the formula-

$$X \text{ ng of PCR product to ligate} = (\gamma \text{ bp of PCR product}) - (50 \text{ ng of vector}) / (\text{size in bp of the vector})$$
Three time “X”ng would be used for a 1:3 molar ratio.
4. Ligation Reaction Mixture: Ligation reaction mixture has been listed in Table 19.2.
5. Incubate at 22 °C for 1 h.
6. For maximum yield, the reaction time can be extended overnight.

19.3.12. Transformation and Selection of the Clones

1. Follow the procedure mentioned in the above sections.
2. Plasmid isolation may be done following the protocol of this manual.

19.3.13. Release of the Insert

To release the insert:

Plasmid DNA (1–2 µg)	–5 µl
10× RE buffer	–2 µl
Restriction enzyme 1 (e.g., <i>EcoRI</i>)	–2 µl
Restriction enzyme 2 (e.g., <i>HindIII</i>)	–2 µl
Deionized distilled water	–20 µl

19.3.14. Sequencing of PCR Product

Sequencing may be done following protocol of “Applied Biosystems BigDye Terminator v3.1 Cycle Sequencing kit” [1].

19.3.15. Removal of Excess of Salts and Enzymes from PCR Products

The PCR amplified products can be subjected for the removal of excess salts as follows:

1. To the template DNA, add enough MQ water to make the volume to 100 µl. Add 10 µl of 3 M sodium acetate pH 5.5 and 250 µl of chilled absolute ethanol.
2. Mix the contents well and incubate on ice for 20–30 min.
Note: Incubation at lower temperatures and for longer periods may cause precipitation of salts and hence is not recommended.
3. Spin at 12,000×*g* for 20 min and remove the supernatant.
4. Wash the pellet by adding 500 µl of 70 % ethanol at room temperature and centrifuge at 12,000×*g* for 5 min.
5. Aspirate or decant the supernatant and repeat the 70 % ethanol wash step once more.
6. Air dry the pellet and resuspend in a suitable volume of water.
7. Check an aliquot on gel and quantification.

19.3.16. Exo–SAP Digestion of PCR Product

1. Make a master mix of *Exonuclease I* and *Shrimp Alkaline Phosphatase* (SAP) for 10 µl of PCR product as per Table 19.3.
2. Add 1 µl of the mastermix to 10 µl of PCR product and set up the following incubation protocol in a thermal cycler:

Hold: 37 °C 120 min
Hold: 85 °C 15 min
Hold: 4 °C

Both *ExoI* and SAP are active in 1× Taq Buffer and can be easily denatured by heating to 85 °C for 15 min.

19.3.17. Sequencing PCR

1. Make a master mix of your sequencing reaction based on the following volumes given in Table 19.4.

Table 19.3
Mastermix of Exo-SAP digestion

Components	Units/Rxn.	1×	100×
<i>Exo</i> I (20 U/μl)	0.5	0.025	2.5
SAP (1 U/μl)	0.5	0.5	50
PCR Buffer 10×	1×	0.1	10
MilliQ		0.375	37.5
Total volume		1	100

Table 19.4
Ingredients of master mix of sequencing reaction

Reaction	1/4×	1/8×
BDT	1	0.5
5× Seq Buffer	1.5	1.75
Primer 3.2 μM	3.2 pmol	3.2 pmol (1 μl)
Template	—	—
Water	To 10 μl	To 10 μl

2. Thaw out your primer first and add the correct amount of this and water to a tube.
3. Thaw out the 5× buffer, mix well, and add the correct amount to the tube.
4. Remove an aliquot of BDT and thaw on ice. Mix well and spin down the tube.
5. Add the correct amount to the reaction mix.
6. Mix the master mix well by inversion and spin down. The master mix is now ready to be aliquoted into strip tubes, a plate, or single tubes.
7. Add the purified template (up to 6.75 μl), typically 1–5 μl for 300–1,500 bp products depending on concentration, based on 3–10 ng for 200–500 bp, 5–20 ng for 500–1,000 bp, or 10–40 ng for 1,000–2,000 bp.
8. Seal the plate with PCR film, or tubes as per normal. Mix the reaction by vortexing for 3 s. Flick the product back down to the bottom of the wells.
9. Place the plate/tubes in a PCR machine.

The reaction is as follows:

96 °C 10 s
50 °C 10 s
60 °C variable ^a
Repeat for 25 cycles
Hold at 15 °C

^aThis is the extension step and so alter the time to be the same as what you would use for a PCR, i.e., $\leq 1,000$ bp/min

- 10. The primer temp can be altered for difficult templates.
- 11. Once the sequencing reaction is finished, the samples can be stored at -20°C for few days or else continue for the cleanup step.

19.3.18. PCR Product Cleanup

After sequencing PCR, the amplified products should be carried out for the removal of the excess salts, primers, and enzymes as follows:

- 1. Transfer the reaction product into a 1.5 ml tube.
- 2. Make a master mix I of 10 μl Milli-Q and 2 μl of 125 mM EDTA per reaction.
- 3. Add 12 μl of master mix I to each reaction containing 10 μl of reaction.
- 4. Ensure the contents are mixed.
- 5. Make master mix II of 2 μl of 3 M NaOAc pH 4.6 and 50 μl of ethanol per reaction.
- 6. Add 52 μl of master mix II to each reaction.
- 7. Mix the contents well and incubate at room temperature for 15 min.
- 8. Spin at a speed of $12,000\times g$ for 20 min at room temperature.
- 9. Decant the supernatant.
- 10. Add 250 μl of 70 % and spin at $12,000\times g$ for 10 min at room temperature.
- 11. Decant the supernatant.

Add 12–15 μl of Hi-Di formamide, transfer to sample tubes cover with septa, denature, snap chill, and proceed for electrophoresis.

19.3.19. Sequence Analysis

The resulting ITS sequences were analyzed for homologies to sequences deposited in the GenBank and EMBL databases.

Reference

1. Applied Biosystems BigDye Terminator v3.1 Cycle Sequencing Kit Protocol.

Chapter 20

Biological Sequence Analysis: Algorithms and Statistical Methods

Suchi Smita, Krishna P. Singh, Bashir A. Akhoon,
and Shailendra K. Gupta

Abstract

With the increase in huge amount of biological sequence data from large genome and proteome sequencing projects, efforts have been made to develop computational algorithms and databases to manage the information. This chapter is an attempt to highlight some of the commonly used algorithms for the biological sequence analysis ranging from pairwise sequence analysis, multiple sequence analysis, phylogenetic analysis, and prediction of the probability of a desired motif in the sequence. The chapter is organized in the form of basic questions that arise in the researchers' mind and their step-by-step solution using important algorithms and statistical methods. The examples are used and elaborated in such a way that the algorithms can be easily understood by students with nonmathematical and nonstatistical background.

20.1 Introduction

A large number of computational algorithms and tools are available to decipher the information hidden in genome and proteome of any living organism. Alignment of the nucleotide and protein sequences is one of the basic protocols, in any of the genomics, proteomics, transcriptomics, and metabolomics project. The knowledge generated out of sequence alignment can be used directly or indirectly in a variety of applications in molecular biology.

20.2 Materials

20.2.1. Pairwise Sequence Alignment Using Dynamic Programming

Dynamic programming (DP) is the method to solve the complex problem by breaking it down to smaller subproblems. In case of any sequencing project, it is crucial to identify the functional annotation of the given sequence which can be achieved by aligning the sequence with a completely annotated sequence database. Considering the size of the nucleotide/protein sequences, the

alignment is not a straightforward task. In case of sequence alignment, DP is the most common algorithm which is used for approximating the matching of two sequences. The individual positions of the sequences are scored by first using a scoring matrix and then followed by a trace back to come to the optimal alignment solution.

Three important steps involved in dynamic programming are as follows:

1. Construction of matrix based on sequence size and setting up of initial conditions
2. Tabular computation (matrix fill)
3. Trace back through matrix for optimal alignment solution

20.2.1.1. Performing Global/Local Pairwise Alignment for Given Set of Sequences

For the given two small sequences below, we will identify global and local alignment between the sequences using Needleman–Wunsch/Smith–Waterman dynamic programming algorithm approach [1].

Sequence 1	AGTAGCTTCCAAA
Sequence 2	AGAGCTCACAA
Score	Match = +4; Mismatch = -4; Gap penalty = -2

20.2.1.2. Construction and Initialization of Matrix

DP for the sequence alignment starts with the construction of a matrix based on sequence length. For the above sequences:

Length of sequence 1 (m) = 13 and

Length of sequence 2 (n) = 11

A matrix of $[(13 + 1) \times (11 + 1)]$ needs to be constructed in this case as given in Fig. 20.1. The first row and first column of the matrix have to be initialized as multiple of gap penalty (-2).

20.2.1.3. Matrix Filling

All the cells of the matrix are scored based on the following formula:

$$M_{ij} = \max \begin{bmatrix} M_{i-1,j-1} + S_{ij} \\ M_{i-1,j} + W \\ M_{i,j-1} + W \end{bmatrix}, \quad (\text{Eq. 20.1})$$

where M_{ij} = maximum score set in the matrix at position i, j , S_{ij} = Match/mismatch score, W = Gap penalty.

To fill the value in the cell (2,2) based on (Eq. 20.1) (cell filled with gray color in Fig. 20.1):

$$M_{ij} = \max \begin{bmatrix} 0 + 4 \\ -2 + -2 \\ -2 + -2 \end{bmatrix} = 4, \text{ as shown in Fig. 20.2.}$$

		A	G	A	G	C	T	C	A	C	A	A
	0	-2	-4	-6	-8	-10	-12	-14	-16	-18	-20	-22
A	-2											
G	-4											
T	-6											
A	-8											
G	-10											
C	-12											
T	-14											
T	-16											
C	-18											
C	-20											
A	-22											
A	-24											
A	-26											

Fig. 20.1 Matrix construction and initialization for the dynamic programming.

		A	G	A	G	C	T	C	A	C	A	A
	0	-2	-4	-6	-8	-10	-12	-14	-16	-18	-20	-22
A	-2	4										
G	-4											
T	-6											
A	-8											
G	-10											
C	-12											
T	-14											
T	-16											
C	-18											
C	-20											
A	-22											
A	-24											
A	-26											

Fig. 20.2 Matrix with the value calculated for the cell (2,2).

		A	G	A	G	C	T	C	A	C	A	A
	0	-2	-4	-6	-8	-10	-12	-14	-16	-18	-20	-22
A	-2	4	2	0	-2	-4	-6	-8	-10	-12	-14	-16
G	-4	2	8	6	4	2	0	-2	-4	-6	-8	-10
T	-6	0	6	4	2	0	6	4	2	0	-2	-4
A	-8	-2	4	10	8	6	4	2	8	6	4	2
G	-10	0	2	8	14	12	10	8	6	4	2	0
C	-12	-6	0	6	12	18	16	14	12	10	8	6
T	-14	-8	-2	4	10	16	22	20	18	16	14	12
T	-16	-10	-4	2	8	14	20	18	16	14	12	10
C	-18	-8	-6	0	6	12	18	24	22	20	18	16
C	-20	-10	-8	-2	4	10	16	22	20	26	24	22
A	-22	-12	-10	-4	2	8	14	20	26	24	30	28
A	-24	-14	-12	-6	0	6	12	18	24	22	28	34
A	-26	-16	-14	-8	-2	4	10	16	22	20	26	32

Fig. 20.3 Matrix with the calculated $M_{i,j}$ value for each cell.

All the cells of the matrix need to be filled in the same procedure. Figure 20.3 has all the matrix cells with the calculated value of $M_{i,j}$ based on (Eq. 20.1).

20.2.1.4. Trace Back for Optimal Alignment

The trace back step determines the actual alignment(s) that result in the maximum score. There are likely to be multiple maximal alignments possible in the given sequences. Trace back starts from the last cell in the matrix. The alignment between the sequences is generated in the reverse order as we move from the last cell to the first cell in the matrix.

For trace back, the reverse arrow is generated pointing towards the cell with maximum value. The direction of the arrowhead determines the gaps in the aligned sequences. A total of three moves of arrow positions are possible:

1. Arrow pointing towards diagonal cell: match/mismatch at the given position
2. Arrow pointing towards upper cell: gap in the upper sequence
3. Arrow pointing towards left cell: gap in the left sequence

The optimal global alignment generated, based on the direction of arrowhead in Fig. 20.4, between two given sequences is as follows:

		A	G	A	G	C	T	C	A	C	A	A
	0	-2	-4	-6	-8	-10	-12	-14	-16	-18	-20	-22
A	-2	4	2	0	-2	-4	-6	-8	-10	-12	-14	-16
G	-4	2	8	6	4	2	0	-2	-4	-6	-8	-10
T	-6	0	6	4	2	0	6	4	2	0	-2	-4
A	-8	-2	4	10	8	6	4	2	8	6	4	2
G	-10	0	2	8	14	12	10	8	6	4	2	0
C	-12	-6	0	6	12	18	16	14	12	10	8	6
T	-14	-8	-2	4	10	16	22	20	18	16	14	12
T	-16	-10	-4	2	8	14	20	18	16	14	12	10
C	-18	-8	-6	0	6	12	18	24	22	20	18	16
C	-20	-10	-8	-2	4	10	16	22	20	26	24	22
A	-22	-12	-10	-4	2	8	14	20	26	24	30	28
A	-24	-14	-12	-6	0	6	12	18	24	22	28	34
A	-26	-16	-14	-8	-2	4	10	16	22	20	26	32

Fig. 20.4 Trace back steps. The arrow indicates the optimal alignment path.

Sequence 1	A	G	-	A	G	C	T	-	C	A	C	A	A	-
Sequence 2	A	G	T	A	G	C	T	T	C	-	C	A	A	A

The optimal global alignment score = 32.

A local alignment was defined as the problem of finding the best alignment between substrings of both sequences. Smith and Waterman in 1981 showed that a local alignment can be computed using essentially the same idea employed by Needleman and Wunsch by slight modification in the calculation of matrix cell value [2]. The formula for computing the value is given below:

$$M_{ij} = \max \begin{bmatrix} M_{i-1,j-1} + S_{ij} \\ M_{i-1,j} + W \\ M_{i,j-1} + W \\ 0 \end{bmatrix} \quad (\text{Eq. 20.2})$$

From the above equation, it is clear that the value of any of the cell in the matrix will always be ≥ 0 . Based on the Smith and Waterman assumption, the first row and the column of the matrix are initialized with zero. The alignment matrix (Fig. 20.4) is

		A	G	A	G	C	T	C	A	C	A	A
	0	0	0	0	0	0	0	0	0	0	0	0
A	0	4	2	4	2	0	0	0	4	2	4	2
G	0	2	8	6	8	6	4	2	2	0	2	0
T	0	0	6	4	6	4	10	8	6	4	2	0
A	0	4	4	10	8	6	8	6	12	10	8	6
G	0	2	8	8	14	12	10	8	10	8	6	4
C	0	0	6	6	12	18	16	14	12	14	12	10
T	0	0	4	4	10	16	22	20	18	16	14	12
T	0	0	2	2	8	14	20	18	16	14	12	10
C	0	0	0	0	6	12	18	24	22	20	18	16
C	0	0	0	0	4	10	16	22	20	26	24	22
A	0	4	2	4	2	8	14	20	26	24	30	28
A	0	4	2	6	4	6	12	18	24	22	28	34
A	0	4	2	6	4	4	10	16	22	20	26	32

Fig. 20.5 Matrix with the calculated $M_{i,j}$ value for each cell using Smith and Waterman algorithm.

		A	G	A	G	C	T	C	A	C	A	A
	0	0	0	0	0	0	0	0	0	0	0	0
A	0	4	2	4	2	0	0	0	4	2	4	2
G	0	2	8	6	8	6	4	2	2	0	2	0
T	0	0	6	4	6	4	10	8	6	4	2	0
A	0	4	4	10	8	6	8	6	12	10	8	6
G	0	2	8	8	14	12	10	8	10	8	6	4
C	0	0	6	6	12	18	16	14	12	14	12	10
T	0	0	4	4	10	16	22	20	18	16	14	12
T	0	0	2	2	8	14	20	18	16	14	12	10
C	0	0	0	0	6	12	18	24	22	20	18	16
C	0	0	0	0	4	10	16	22	20	26	24	22
A	0	4	2	4	2	8	14	20	26	24	30	28
A	0	4	2	6	4	6	12	18	24	22	28	34
A	0	4	2	6	4	4	10	16	22	20	26	32

Fig. 20.6 Matrix with the optimal trace back path for the best local alignment between two sequences.

recalculated based on (Eq. 20.2) to obtain the local alignment between the given sequences as shown in Fig. 20.5.

Another important distinction is that the score of the best local alignment is the highest value found anywhere in the matrix. This position will be the starting point for trace back to retrieve an optimal alignment using the same procedure described for the global alignment case as shown in Fig. 20.6. However, the path

ends as soon as an entry with score zero is reached. It is trivial to see that the Smith–Waterman algorithm has the same time and space complexity as the Needleman–Wunsch algorithm.

The local alignment between two sequences given in the problem based on Smith and Waterman assumption can be identified as

Sequence	A	G	–	A	G	C	T	–	C	A	C	A	A
Sequence a	A	G	T	A	G	C	T	T	C	–	C	A	A

The best local alignment score = 34.

20.2.2. Multiple Sequence Alignment

Multiple sequence alignment is applied to a set of sequences that are assumed to be homologous (have a common ancestor sequence) with the goal to detect homologous residues and place them in the same column of the multiple alignment. To calculate evolutionary homology, multiple alignments are better suited than pairwise alignments, since if several sequences are compared simultaneously, the chance of random similarities occurring becomes much lower. Therefore multiple alignments can be used both for similarity as well as dissimilarity as in case of classify members of protein families and for phylogenetic analysis. Multiple alignment is also important for computing profiles, prediction of protein secondary structure, or computation of sequence motifs, etc.

Assuming that we have established a family F of homologous protein sequences,

where $F = \{A_1; A_2; : : : ; A_r\}$,

now multiple sequence alignment can be used to predict if a new sequence A_0 belong to the family F or not. One method is to align A_0 to each of $A_1; A_2; : : : ; A_r$ in turn. If one of these alignments produces a high score, then we may decide that A_0 belongs to the family F . However, perhaps A_0 does not align particularly well to any one specific family member, but may score well in a multiple alignment.

Suppose we are given r sequences $A_i; i = 1; : : : ; r$ over an alphabet Σ :

$$A := \begin{cases} A_1 = (a_{11}, a_{12}, \dots, a_{1n_1}) \\ A_2 = (a_{21}, a_{22}, \dots, a_{2n_2}) \\ \vdots \\ A_r = (a_{r1}, a_{r2}, \dots, a_{rn_r}) \end{cases}.$$

A multiple sequence alignment (msa) of A is obtained by inserting gaps (“–”) into the original sequences such that all resulting sequences A_i^* have equal length $L \geq \max \{n_i | i = 1, \dots, r\}$,

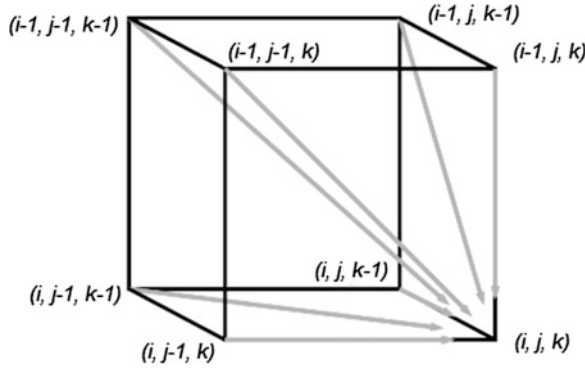
$A_i^* = A_i$ after removal of all gaps from A_i^* , and no column consists of gaps only:

$$A^* := \begin{cases} A_1^* = (a_{11}^*, a_{12}^*, \dots, a_{1L}^*) \\ A_2^* = (a_{21}^*, a_{22}^*, \dots, a_{2L}^*) \\ \vdots \\ A_r^* = (a_{r1}^*, a_{r2}^*, \dots, a_{rL}^*) \end{cases}$$

In case of multiple sequence alignment, Needleman–Wunsch Algorithm could be generalized using scoring function as the sum of column scores:

$$S(m) = \sum_i S(m_i).$$

Consider the alignment with three sequences; a three-dimensional matrix has to be generated as in case of two-dimensional matrix in pairwise alignment. Each cell in the matrix has at most seven neighboring cells and score for each cell in the matrix can be calculated as



$$S_{i,j,k} = \max \begin{cases} S_{i-1,j-1,k-1} + \delta(v_i, w_j, u_k) \\ S_{i-1,j-1,k} + \delta(v_i, w_j, -) \\ S_{i-1,j,k-1} + \delta(v_i, -, u_k) \\ S_{i,j-1,k-1} + \delta(-, w_j, u_k) \\ S_{i-1,j,k} + \delta(v_i, -, -) \\ S_{i,j-1,k} + \delta(-, w_j, -) \\ S_{i,j,k-1} + \delta(-, -, u_k) \end{cases}$$

$\delta(x, y, z)$ is an entry in the 3-D scoring matrix

For three sequences of length n , the run time is $7n^3!O(n^3)$, and for k sequences, total run time can be calculated as $(2^k - 1)(n^k)!O(2^k n^k)$.

In case of multiple alignment, dynamic programming approach for alignment between two sequences is easily extended to k sequences, but it is impractical due to exponential running

time. For the given multiple sequence alignment between three sequences:

x: AC-GCGG-C
y: AC-GC-GAG
z: GCCGC-GAG

Following three pairwise alignments may be induced:

Pair 1	x: A C G C G G – C
	y: A C G C – G A C
Pair 2	x: A C – G C G G – C
	z: G C C G C – G A G
Pair 3	y: A C – G C G A G
	z: G C C G C G A G

Thus from any given multiple alignment, we can infer pairwise alignments between all the sequences, but they may not be necessarily optimal. However, in many situations, it is difficult to infer a “good” multiple alignment from optimal pairwise alignments between all sequences.

Since exact methods of multiple sequence alignment have exponential time complexity, heuristic approaches such as Progressive Alignments are generally used which is based on the profile. Generation of profile for multiple alignments is a very rich representation of alignment. In a profile of multiple alignment, we can show the information in each column and even the pairwise relationships in a profile. Here we can identify the relative frequencies of the alphabets and Gaps. Also we can optionally identify the different kinds of Gap events like Gap Open, Gap Close, and Gap Extension.

Below are the multiple alignments between five sequences and the profile generated with the information available in the multiple alignments.

MSA	–	A	G	G	C	T	A	T	C	A	C	C	T	G
	T	A	G	–	C	T	A	C	C	A	–	–	–	G
	C	A	G	–	C	T	A	C	C	A	–	–	–	G
	C	A	G	–	C	T	A	T	C	A	C	–	G	G
	C	A	G	–	C	T	A	T	C	G	C	–	G	G
A		1					1		0.8					
C	0.6				1		0.4	1		0.6	0.2			
G		1	0.2						0.2			0.4	1	

(continued)

T	0.2	1	0.6	0.2
O	0.2	0.8	0.4	0.4
E				0.4
C	0.2	0.8	0.4	0.4

where A, C, G, T are nucleotides.
O = Gap Open is defined to be a gap where previous character in the sequence is not a gap.
C = Gap Close is defined to be a gap where next character in the sequence is not a gap.
E = Gap Extension is defined to be a gap where both next and previous characters are gap.
Profile representation is important to align sequence to sequence, sequence to a given profile, and profile to profile alignment. For the given two alignments:

x: GGGCACTGCAT	Alignment 1
y: GGTTCAGTC --	
z: GGGAAGTCAG	
w: GGACGTACC --	Alignment 2
v: GGACCT - - - -	

Both the alignment may be merged into a single multiple alignments by creating their corresponding profiles as below:

x: GGGCACTGCAT	Combined alignment
y: GGTTCAGTC --	
z: GGGAAGTCAG	
w: GGACGTACC --	
v: GGACCT - - - -	

20.2.2.1. Greedy Approach for Multiple Alignment

For the given k sequences, the Greedy approach for multiple alignment is to choose the most similar pair of strings and combine them into a profile, thereby reducing alignment of k sequences to an alignment of $k-1$ sequences/profiles [3]. The step will be repeated until all the sequences are aligned. Below is the example of greedy approach:

$u_1 = \text{ACGTACGTACGT}$

$u_2 = \text{TTAATTAATTAA}$

$u_3 = \text{ACTACTACTACT}$

\vdots

$u_k = \text{CCGGCCGGCCGG}$

$\left. \begin{array}{l} \vdots \\ \vdots \\ \vdots \end{array} \right\} k-1$

$u_1 = \text{AG}_{g/t}\text{TAC}_{g/t}\text{TAC}_{g/c}\text{T} \dots$

$u_2 = \text{TTAATTAATTAA} \dots \dots$

\dots

$u_k = \text{CCGGCCGGCCGG} \dots$

In the example above, sequences u_1 and u_3 are merged as a profile because they represent the most similar pair of strings in the example. The more realistic example for the greedy approach is discussed below:

20.2.3. To Perform Multiple Sequence Alignment for Given Set of Sequences Using Greedy Approach

Consider these four sequences

s1 = GATTCA
s2 = GTCTGA
s3 = GATATT
s4 = GTCAGC

Score: match 1, mismatch, and indels −1
Total number of possible alignment pair can be calculated as

$$\text{Possible alignment} = \frac{n!}{(n-2)! * 2!}$$

where n = total no. of sequences.

20.2.3.1. To Perform the Pairwise Alignment Between All the Four Sequences as Discussed in Protocol 1

Below is the pairwise alignment score between all the possible pairs:

	Pairwise alignment	Alignment score
Pair 1	s1 G A T – T C A	1
	s2 G – T C T G A	
Pair 2	s1 G A T – T C A	1
	s3 G A T A T – T	
Pair 3	s1 G A T T C A – –	0
	s4 G – T – C A G C	
Pair 4	s2 G – T C T G A	−1
	s3 G A T A T – T	

(continued)

	Pairwise alignment	Alignment score
Pair 5	s2 G T C T G A	2
	s4 G T C A G C	
Pair 6	s3 G A T – A T T	–1
	s4 G – T C A G C	

From the above pairwise alignment, Pair 5, representing sequences 2 and 4, has the maximum alignment score based on the scoring parameters given.

20.2.3.2. Combining of the Two Sequences

For multiple alignment using greedy approach, we can combine these two sequences as below:

$$\begin{array}{l} \text{s2 G T C T G A} \\ \text{s4 G T C A G C} \end{array} \longrightarrow \text{s2,4 G T C } \overset{\text{t}}{\underset{\text{a}}{\text{c}}} \text{G } \overset{\text{a}}{\underset{\text{c}}{\text{c}}}.$$

Now we have a set of three sequences:

s1 = GATTCA
s3 = GATATT
s2,4 = GTCt/aGa/c

20.2.3.3. Designing of Pairs and Identification of Best Alignment

We have to repeat the process for designing pairs and identification of best alignment between the pairs. Below is the optimal pairwise alignment between three sequences (s1, s3, and s2,4)

	Pairwise alignment	Alignment score
Pair 1	s1 G A T – T C A	1
	s3 G A T A T – T	
Pair 2	s1 G A T T C – – A	0
	s2,4 G – T – C T G A	
Pair 3	s3 G A T A T T –	–1
	s2,4 G – T C T G A	

From the next set of pairwise alignment, pair s1 and s3 has the maximum alignment score.

20.2.3.4. Combining of Two Sequences to Get a New Profile

We can combine sequences 1 and 3 to create a profile of s1,3.

$$\begin{array}{l} \text{s1 G A T – T C A} \\ \text{s3 G A T A T – T} \end{array} \longrightarrow \text{s1,3 G A T A T C } \overset{\text{a}}{\underset{\text{t}}{\text{t}}}.$$

Finally, we have a set of two sequences:

s1,3 = G A T A T C a/t
s2,4 = G T C t/a G a/c

which can be aligned as below with the alignment score of 1 based on given parameters:

s1,3	G A T A T C A
s2,4	G – T C T G A

20.2.3.5. Final Multiple Sequence Alignment by Greedy Approach

Based on greedy approach, the final multiple sequence alignment between the four sequences can be generated by considering a rule that “once a gap, always a gap.”

s1 G A T – T C A
s2 G A T A T – A
s3 G – T C T G A
s4 G – T C A G C

20.2.3.6. Progressive Alignment Method for Multiple Sequence Alignment

Another variation of greedy algorithm for the multiple sequence alignment is the progressive alignment which has somewhat more intelligent strategy for choosing the order of alignments between the given sequences. Progressive alignment works well for close sequences, but not suitable for the distantly related sequences. Progressive alignment uses profiles to compare sequences and the gaps in consensus string are permanent. It is one of the most common approaches for multiple sequence alignment. In general, this works by constructing a series of pairwise alignments, first starting with pairs of sequences and then followed by aligning sequences to existing alignments (profiles) and profiles to profiles.

Progressive alignment is a heuristic and does not directly optimize any known global scoring function of alignment correctness. However, it is fast and efficient, and often provides reasonable results [4]. The progressive alignment results differ with the order in which the sequences are aligned. Also, generation of single or several profiles influence the result. Result may also vary with the scoring function used. Most of the multiple alignment tools are based on either the complete alignment or the pair-guided alignment method.

In case of complete alignment, all sequences, from two sub-alignments, are used in a dynamic programming approach where sequences, from one subset, are aligned in a pairwise manner with all the sequences in another aligned subset. Alignment with the maximum score will determine how the sequence will be aligned to the group. Another approach is the pair-guided alignment, where two specific sequences are chosen from each group of sub-alignment. The alignment between these two sequences determines the final alignment between the groups.

The most crucial step of the heuristic of progressive alignment algorithms is to determine the order of sequences to align first.

Clearly the most consistent alignments are those which align very similar pairs of sequences first. To determine which sequences are similar, most of the algorithms build a guide tree. The guide tree is a binary phylogenetic alignment tree, where the root node represents the multiple alignment. The nodes furthest away from the root are the most similar pairs. The methods used to compute the guide tree are similar to the distance-based methods used for reconstructing phylogenetic trees, though they are often the rather “quick and dirty” ones. Feng and Doolittle in 1987 published the first progressive alignment algorithm which is based on the idea that a pair of sequences with minimal distance has also evolutionary diverged most recently [4]. Thus, for the optimal multiple sequence alignment, one has to follow the evolutionary path of the sequence.

The Feng and Doolittle’s progressive multiple alignment approach is based on first the construction of all the pairwise alignment pair between the set of given sequences and calculation of the alignment scores which is then converted into a distance score. Based on the distance score, a distance matrix is generated that helps to construct a guide tree. In the guide tree, the approach is to start from the first node that was added to the tree and align the two children nodes (which may be two sequences, one sequence and one sub-alignment, or two sub-alignments). These steps are repeated for all other nodes in their tree order until one reaches the root, i.e., until all sequences have been aligned [5].

The distance between the alignment pair is the normalized percentage similarity, which is calculated as

$$\text{Similarity} = \frac{\text{Exact match in the alignment pair}}{\text{Total length of the alignment}}.$$

20.2.4. Performing Multiple Sequence Alignment for Given Set of Sequences Using Progressive Alignment Method

Consider the following five small sequences:

s1 = ATTGCCATT
s2 = ATGGCCATT
s3 = ATCCAATTTT
s4 = ATCTTCTT
s5 = ATTGCCGATT

20.2.4.1. Optimal Pairwise Alignment

Randomly select one sequence as the base sequence. Perform the optimal pairwise alignment with other sequences in the group

For example, sequence s1 is selected as the base sequence.

	Pairwise alignment
Pair 1	s1 ATTGCCATT
	s2 ATGGCCATT
Pair 2	s1 ATTGCCATT – –
	s3 ATC – CAATTTT

(continued)

Pair 3	s1 ATTGCCATT
	s4 ATCTTC – TT
Pair 4	s1 ATTGCC – ATT
	s5 ATTGCCGATT

20.2.4.2. *Once a Gap
Always a Gap*

Merge the pairwise alignment using the rule “once a gap, always a gap” as shown below:
Merging pairwise alignment of pair 1, 2

s1	A	T	T	G	C	C	A	T	T	–	–
s2	A	T	G	G	C	C	A	T	T	–	–
s3	A	T	C	–	C	A	A	T	T	T	T

Merging pair 3 alignment with the above alignment

s1	A	T	T	G	C	C	A	T	T	–	–
s2	A	T	G	G	C	C	A	T	T	–	–
s3	A	T	C	–	C	A	A	T	T	T	T
s4	A	T	C	T	T	C	–	T	T	–	–

Finally, merging pair 4 alignment with the above alignment. As alignment pair 4 has gap in s1, the entire column will be shifted to incorporate gap.

s1	A	T	T	G	C	C	–	A	T	T	–	–
s2	A	T	G	G	C	C	–	A	T	T	–	–
s3	A	T	C	–	C	A	–	A	T	T	T	T
s4	A	T	C	T	T	C	–	–	T	T	–	–
s5	A	T	T	G	C	C	G	A	T	T	–	–

20.2.4.3. *To Obtain Best
Alignment*

Choose another sequence as a base sequence and repeat step 1 and 2 to return the best multiple alignment.

20.2.4.4. *Construction of
Phylogenetic Tree Using
UPGMA Method*

Unweighted Pair Group Method with Arithmetic Mean is a simple hierarchical clustering method used for the creation of phylogenetic trees. UPGMA assumes a constant rate of evolution, and is not a well-regarded method for inferring phylogenetic trees unless this assumption has been tested and justified for the dataset being used. The algorithm iteratively joins the two nearest clusters, until only one cluster is left.

20.2.4.5. UPGMA Algorithm

Let d be the distance function between species; we define the distance $D_{i,j}$ between two clusters of species C_i and C_j as follows:

$$D_{i,j} = \frac{1}{n_i + n_j} \sum_{p \in C_i} \sum_{q \in C_j} d(p, q)$$

where $n_i = |C_i|$ and $n_j = |C_j|$

20.2.4.6. Initialization of Algorithm

1. Initialize n clusters with the given species, one species per cluster
2. Set the size of each cluster to 1: $n_i \leftarrow 1$
3. In the output tree T , assign a leaf for each species

20.2.4.7. Iteration

1. Find the i and j that have the smallest distance D_{ij} .
2. Create a new cluster— (ij) , which has $n_{(ij)} = n_i + n_j$ members.
3. Connect i and j on the tree to a new node, which corresponds to the new cluster (ij) , and give the two branches connecting i and j to (ij) length $\frac{D_{ij}}{2}$ each.
4. Compute the distance from the new cluster to all other clusters (except for i and j , which are no longer relevant) as a weighted average of the distances from its components:

$$D_{(ij),k} = \left(\frac{n_i}{n_i + n_j} \right) D_{i,k} + \left(\frac{n_j}{n_i + n_j} \right) D_{j,k}$$

5. Delete the columns and rows in D that correspond to clusters i and j , and add a column and row for cluster (ij) , with $D_{(ij),k}$ computed as above.
6. Return to 1 until there is only one cluster left.

20.2.5. Construction of Phylogenetic Tree Using Neighbor Join Algorithm

Neighbor joining is a bottom-up clustering method used for the construction of phylogenetic trees. The neighbor joining method is a greedy heuristic which joins at each step, the two closest subtrees that are not already joined. It is based on the minimum evolution principle. One of the important concepts in the NJ method is *neighbors*, which are defined as two taxa that are connected by a single node in an unrooted tree [6].

20.2.5.1. Initialization of Algorithm

Same as initialization process of UPGMA

20.2.5.2. Iteration

For each species, compute $u_i = \sum_{k \neq i} \frac{D_{i,k}}{(n-2)}$

where u_i is the distance of node i from rest of the tree

1. Choose the i and j for which $D_{i,j} - u_i - u_j$ is smallest.
2. Join clusters i and j to a new cluster (ij) , with a corresponding node in T . Calculate the branch lengths from i and j to the new node as

$$d_{i,(ij)} = \frac{1}{2} D_{i,j} + \frac{1}{2} (u_i - u_j), d_{j,(ij)} = \frac{1}{2} D_{i,j} + \frac{1}{2} (u_j - u_i)$$

3. Compute the distances between the new cluster and each other cluster:

$$D_{(ij),k} = \frac{D_{i,k} + D_{j,k} - D_{i,j}}{2}$$

4. Delete clusters i and j from the tables, and replace them by (ij) .
5. If more than two nodes (clusters) remain, go back to 1. Otherwise, connect the two remaining nodes by a branch of length $D_{i,j}$.

20.2.6. To Construct Phylogenetic Tree Using UPGMA and Neighbor Joining Methods

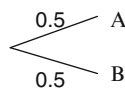
Consider four sequences:

A = ACTA
B = ACTT
C = CGTT
D = AGAT

20.2.6.1. Calculating Number of Mismatches (Distances) Between Two Sequences

	A	B	C
B	1		
C	3	2	
D	3	2	2

A, B is the most similar pair; it will be clustered together as follows:



20.2.6.2. Calculation of New Distance Matrix

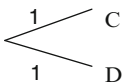
New distance matrix will be calculated as

$$\text{dist}(AB), C = (\text{dist } AC + \text{dist } BC)/2 = 2.5$$

$$\text{dist}(AB), D = (\text{dist } AD + \text{dist } BD)/2 = 2.5$$

	AB	C
C	2.5	
D	2.5	2

Now CD is the most similar pair; it will be clustered together



20.2.6.3. Calculation of Average Distance

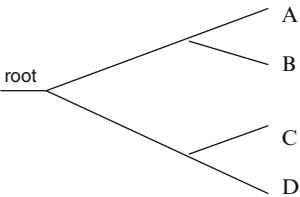
Final calculation is to take the average distance between two composite sets of sequences AB and CD. The average distance will be calculated as

$$\text{dist}(AC + AD + BC + BD)/4 = 2.5$$

	AB
CD	2.5

One-half of this distance $2.5/2 = 1.25$ is included in the part of the tree that goes from the root to CD, and the other half goes from the root to AB.

The best tree generated using UPGMA method



20.2.7. Construction of Phylogenetic Tree Using Neighbor Joining Method

We will generate the neighbor joining tree for the sequence data given in protocol 4.

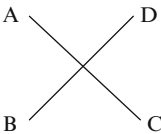
We can define the evolutionary distance between the sequences as follows:

$$d_{ij} = -\ln(1 - p_{ij})$$

where d_{ij} is the distance between sequence i and j and p_{ij} is the fraction of mismatches in the pairwise alignment of sequences i and j .

20.2.7.1. Designing of a Base Tree

Design a base tree as follows:



20.2.7.2. Calculation of the Distance Between All Possible Pairs

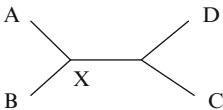
Calculate the distance between all the possible pair of four sequences using the formula:

$$d_{ij} = -\ln(1 - p_{ij})$$

Construction of the distance matrix

	A	B	C	D
A	0			
B	0.28768	0		
C	1.38629	0.69314	0	
D	1.38629	0.69314	0.69314	0

From the matrix, A and B have least distance; we will choose them as neighbor and join them as a new node “X.”



20.2.7.3. Calculation of the Distance

Now, we will again calculate the distance

	AB	C	D
AB	0		
C	1.039715	0	
D	1.039715	0.69314	0

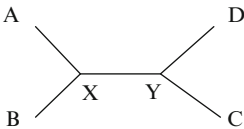
Distance of C from the node X can be calculated as

$$d(C, AB) = [d(C, A) + d(C, B)]/2$$

20.2.7.4. Joining of the Sequence C and D to Create a Complete Tree

We will now join sequence C and D to create a complete tree

The best three generated between the given sequences using neighbor joining method



20.2.8. Markov and Hidden Markov Model

Markov model (MM) and hidden Markov model (HMM) are supervised machine learning techniques that have many applications in bioinformatics such as sequence comparison, gene

finding, homology modeling to identify known folds in a target sequence, phylogeny and functions prediction, etc. MM consists of several states from a system, which represent observations at a specific point in time, and a set of transition probabilities. HMM consists of a set of states just like MM along with hidden states that has probability distribution defined as emission probability.

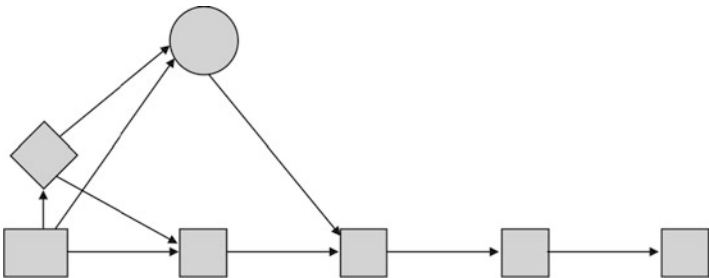
20.2.8.1. HMM for Multiple Sequence Alignment

Generate the profile HMM for multiple sequence alignment.
Consider the multiple alignment of five DNA sequences:

T	C	–	C	T	–
–	C	A	C	T	G
T	–	–	C	T	A
T	G	–	G	C	A
C	C	–	T	T	C

20.2.8.2. Matching (M), Insert (I), and Delete (D) of the Multiple (MSA) Sequence Alignment

In order to generate the profile HMM, we have to define matching (M), insert (I), and delete (D) states of the multiple (MSA) sequence alignment
Column with more than half gap positions are considered as inserts. In the multiple sequence alignment given above, column 3 will be the insert as only in one sequence there is a base, while



column 1, 2, 4, 5, 6 to be considered as the match states. The nucleotide in the match state will be defined, based on consensus generated after multiple alignment.
A sample representation of HMM. The box represents match state, diamond represents the insertion in the sequence, and circle represents deletion

20.2.8.3. Assumptions from the Multiple Sequence Alignment

From the multiple sequence alignment given above, following assumptions can be made:

- From Column 1 to 2 in the MSA
- 3 transition to next match state
- 1 transition to delete state
- 1 transition to insertion state

From Column 2 to 3 in the MSA

- 0 transition to next match state
- 4 transition to delete state
- 1 transition to insertion state

From Column 3 to 4 and 4 to 5 in the MSA

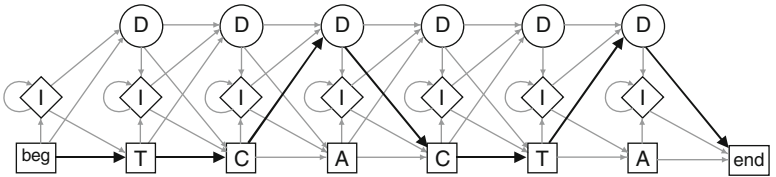
- 5 transition to next match state
- 0 transition to delete state
- 0 transition to insertion state

From Column 5 to 6 in the MSA

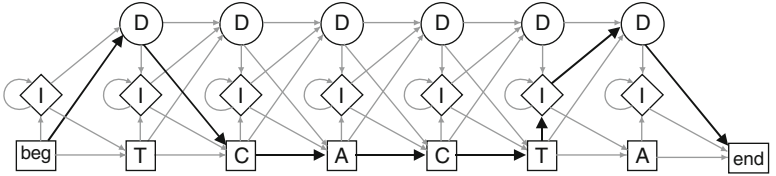
- 4 transition to next match state
- 1 transition to delete state
- 0 transition to insertion state

20.2.8.4. The Hidden Markov Model

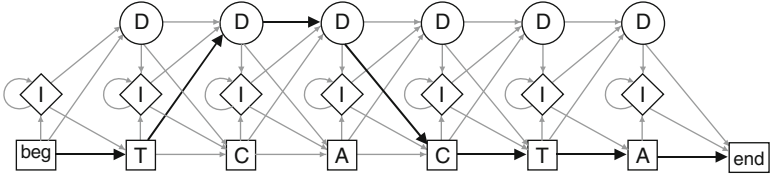
The HMM for all the five sequences in the multiple sequence alignment can be represented as follows:



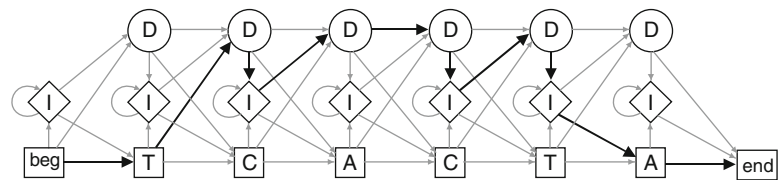
HMM for sequence 1



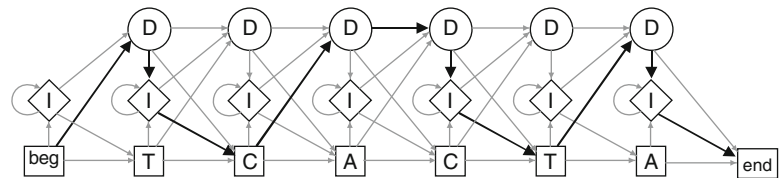
HMM for sequence 2



HMM for sequence 3



HMM for sequence 4



HMM for sequence 5

The transition path between various states in the models is represented as dark black arrow

20.2.8.5. Calculation of the Probability Matrix

From HMM of all the individual sequences, the probability matrix can be calculated as

	1	2	3	4	5	6
A	0	0	1/5	0	0	2/5
T	3/5	0	0	1/5	4/5	0
G	0	1/5	0	1/5	0	1/5
C	1/5	3/5	0	3/5	1/5	1/5
–	1/5	1/5	4/5	0	0	1/5

20.2.8.6. Calculation of Various Parameters of Profile HMM (Match Emissions/Insert Emissions/State Transitions)

		0	1	2	3	4	5
match emissions	A	–	0	0	0	0	2
	C	–	1	3	3	1	1
	G	–	0	1	1	0	1
	T	–	3	0	1	4	0
insert emissions	A	0	0	1	0	0	0
	C	0	0	0	0	0	0
	G	0	0	0	0	0	0
	T	0	0	0	0	0	0
state transitions	M–M	4	3	4	5	4	4
	M–D	1	1	0	0	0	0
	M–I	0	0	0	0	0	0
	I–M	0	0	1	0	0	0
	I–D	0	0	0	0	0	0
	I–I	0	0	0	0	0	0
	D–M	–	1	0	0	0	1
	D–D	–	0	1	0	0	0
	D–I	–	0	0	0	0	0

Based on the probability matrix and the match emission, insert emissions, and state transition parameters, the overall HMM for multiple sequence alignment can be summarized as shown in Fig. 20.7.

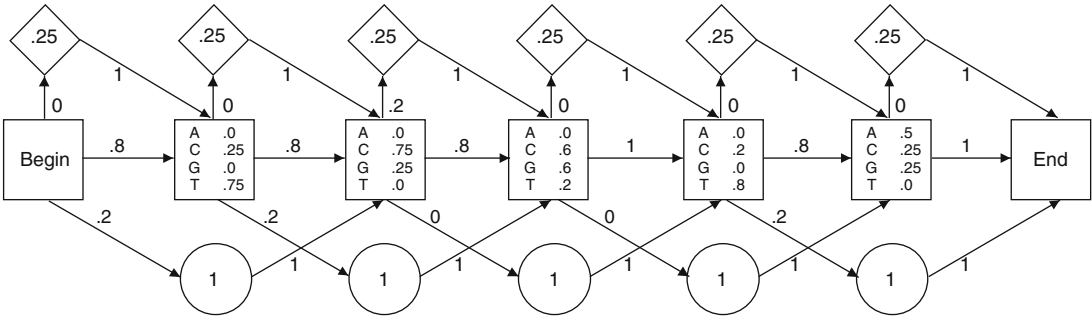


Fig. 20.7 The HMM for the multiple sequence alignment of five sequences. The probability between various transitions states of HMM is mentioned. The square represent match emission, the diamond represents insert emission, and circle represents deletion.

In many cases the Laplace correction has to be applied. Typically Laplace’s rule (add one observation per symbol per column) is used to smooth distributions, along with relative frequency estimation. However, this model does not provide a mechanism to insert symbols between columns, which may be necessary. Such a mechanism is provided by a profile HMM and can be estimated from a given MSA, using relative frequency estimation and simple smoothing.

For the multiple sequence alignment given above, the Laplace correction can be made in the profile as follows:

Base	1	2	3	4	5	6
A	1	1	2	1	1	3
T	4	1	1	2	5	1
G	1	2	1	2	1	2
C	2	4	1	4	2	2
Gap	2	2	5	1	1	2

20.2.8.7. Assigning
Transition Probability

Probability Using Lapalace’s Rule from Column 1 is
4/10 state M1M2
2/10 state M1D2
2/10 state M1I2
2/10 state D1D2

Probability Using Laplace's Rule from Column 2 is
 1/10 state M2M3
 5/10 state M2D3
 2/10 state M2I3
 2/10 state D2D3

Probability of other column will be calculated in a similar fashion

20.2.8.8. Assigning Emission Probability

For the first column there are 3 T's and 1 C and 1 gap. Using Laplace's rule, this will become 4 T; 2 C; 1 A; 1 G and 2 Gaps
 Thus the probabilities will be 4/10 T; 2/10 C; 1/10 A; 1/10 G and 2/10 for Gaps. Probabilities for other column can also assign in similar way.

20.2.9. Markov Model

Generate the probability matrix for coding and noncoding region in a gene using zero- and first-order Markov model. A fundamental Markov model of a process is a model where every state corresponds to an observable event and the state transition probabilities depend just on the current and predecessor state. The Gene finding in DNA has become one of the most important computational biology problems that can be determined using Markov model [7]. A Markov chain is defined by a state space, which is the set of possible values that the Markov chain can take, and the transition probabilities of the process.

20.2.9.1. K-Order Markov Model

$$P(X) = p(x_L, x_{L-1}, \dots, x_1) \\ = p(x_L | x_{L-1}, \dots, x_1) \cdot p(x_{L-1} | x_{L-2}, \dots, x_1) \dots p(x_1).$$

Let as assume to design zero- and first-order Markov models for the coding and noncoding region of mtb48 gene of *Mycobacterium tuberculosis* (GenBank id: AY029285.1) from the position: 207–1,583 and position: 1,591–2,138, respectively.

For the coding sequence of mtb48 gene of *M. tuberculosis* (GenBank id: AY029285.1 position: 207–1,583)

acgcagtcgcagaccgtgacgggtggatcagcaagagatttgaacaggccaacgaggtggaggcc
 ccgatggcggaccacgcactgatgtcccatcacaccgtgcgaactcacggcgctaaaaacgcgcg
 ccaacagctgtattgtccgccgacaacatgcgggaatactggcgccgggtgccaaagagcggcag
 cgtctggcgacctcgtcgcgaacgcggccaaggcgtatggcgaggttgatgaggaggctgcgacc
 gcgctggacaacgacggcggaaggactgtgcaggcagaatcgccggggccgtcggagggggaca
 gttcggccgaactaacgatacgcggagggtggccacggccgtgaaccaactcatggtatctaaa
 gaagcggcaaggaagctcgaacggcgaccaaggcgcatcgctcgcgactttgcggatgggtgg
 aacactttcaacctgacgtcgaaggcgacgtcaagcgggttcgggggttgacaactgggaaggcg
 atgcggctaccgcttgcgaggcttcgctgatcaacaacggcaatggataccacatggccaattga
 gcgctgcgatggccaagcaggctcaatatgtcgcgcagctgcacgtgtgggttagggcggaacatcc
 gacttatgaagacatagtcgggctcgaacggctttacgcggaaaaccccttcggcccgaccaaattct
 cccggtgtacgggagtatcagcagaggtcggagaaggtgctgaccgaataacaacaaggcagcc
 ctggaaccggtaaaccgccgaagcctcccccccatcaagatcgaccgccccgcctccgcaaga
 gcagggttgatccctggcttctgatgcccgctgacggctcgggtgtgactccgggtaccggga

tgccagccgcaccgatggttcgcctaccggatcgccgggtggtggcctccggctgacacggcgg
cgagctgacgtcggctggcggaagccgcagcgtgctggcgacgtggcggtcaaagcggc
atcgctcgggtggcgtggaggcggcggtgctcgccgcttgggatccgcgctcggggc
gccgaatcggtgcggcccctggcgctggtgacattgccgcttaggccagggaaggcgccgg
cggcgccgcgctggcgccgggtggcatgggaatgccgatgggtgccgcgcatcagggaagg
ggcgccaagtccaagggttctcagcaggaagacgaggcgctctacaccgaggtcgggcatggacc
gaggccgtcattggtaccgtcgcgccaggacagtaaggagtccaag

Total number of bases is 1,377 with the base count as follows:
A: 281; C: 410; G: 488; T: 198

20.2.9.2. For Zero-Order
Markov Model

$$P(X) = p(x_L) \cdot p(x_{L-1}) \dots p(x_1),$$

where $x = \{A, G, T, C\}$.
In the zero-order Markov model, the probability distribution of the next base to be generated in the sequence does not depend on any of the base preceding it.
Thus the zero-order probability matrix for the coding region of mtb48 gene of *M. tuberculosis* can be defined as

A	C	G	T
0.204	0.298	0.354	0.143

20.2.9.3. For First-Order
Markov Model

$$P(X) = p(x_L|x_{L-1}) \cdot p(x_{L-1}|x_{L-2}) \dots p(x_2|x_1) \cdot p(x_1).$$

In the first-order Markov model, the probability distribution of the next base to be generated in the sequence depends on the base preceding it.
where $x = \{A, G, T, C\}$.
 $p(x_L|x_{L-1})$ is the probability of observing x_L after x_{L-1} in the sequence. The solution is to obtain the dinucleotide frequency data for the sequence as below:

$P(AA)$	0.27018
$P(AT)$	0.18246
$P(AG)$	0.26667
$P(AC)$	0.2807
$P(TA)$	0.11881
$P(TT)$	0.15347
$P(TG)$	0.40594
$P(TC)$	0.32178
$P(GA)$	0.20367
$P(GT)$	0.12016
$P(GG)$	0.35031

(continued)

$P(GC)$	0.32587
$P(CA)$	0.20048
$P(CT)$	0.14493
$P(CG)$	0.3913
$P(CC)$	0.26329

The probability matrix can be generated based on the above dinucleotide frequency for the given sequence of mtb48 of *M. tuberculosis*.

Probability matrix of the first-order Markov chain model for the coding region of mtb48 gene of *M. tuberculosis*

	A	G	T	C
A	0.270463	0.266904	0.181495	0.281139
G	0.203285	0.351129	0.119097	0.326489
T	0.116162	0.409091	0.151515	0.323232
C	0.200000	0.392683	0.143902	0.263415

For the noncoding sequence of mtb48 gene of *M. tuberculosis* (GenBank id: AY029285.1 position: 1,591–2,138)

ggacgaattggaccgcgatgtcgccggcggttgacgtggcgcgcggttcagtcggccctagacggg
acgtcaatcagatgaacaacggatcctccgcgccaccgacgaagccgagaccgtcgaagtacgatcaat
gggcaccagtggctcaccggcctgcgcacgaagatggtttgcgtgaagaagctgggtgccgagggcggtgg
ctcagcgggtcaacgagggcgtgcacaatgcgcaggccgcggtccgcgtataacgacgcggcgggcg
agcagctgaccgctgcgttatcgccatgtcccgcgcgatgaacgaagggaatggcctaagcccattgttcg
gtggtagcgactacgcaccgaatgagcgcgcgaatgcggtcattcagcgcgccgacacggcggtgagtacg
cattgtcaatgtttgacatggatcgccgggttcggaggcgccatagtcctggtcgccaattatgccgcag
ctagctggtcttaggttcggttacgtggttaattatgacgtccgttacca

Sequence length = 548

Base count

A: 107; C: 156; G: 183; T: 102

Zero-order probability matrix for the noncoding region of the above mentioned sequence can be defined as

A	C	G	T
0.195	0.285	0.334	0.186

**20.2.10. Determination
of the Probability
of DNA Fragment
“AGTAGCTTCCAG”
in the Coding Region
of mtb48 Gene of
M. Tuberculosis Using
Probability Matrix
Generated in Protocol
20.2.9**

For the Sequence $X = \text{AGTAGCTTCCAG}$

**20.2.10.1. Zero-Order
Markov Model Probability
in the Coding Region of
mtb48 Gene**

Zero-order Markov model probability in the coding region of mtb48 gene will be

$$\begin{aligned} P(X) &= P(A) * P(G) * P(T) * P(A) * P(G) * P(C) * P(T) * P(T) \\ &\quad * P(C) * P(C) * P(A) * P(G) \\ &= 0.204 * 0.354 * 0.143 * 0.204 * 0.354 * 0.298 * 0.143 * 0.143 \\ &\quad * 0.298 * 0.298 * 0.204 * 0.354 \\ P &= 2.91445\text{E}-08 \end{aligned}$$

**20.2.10.2. First-Order
Markov Model Probability
in the Coding Region of
mtb48 Gene**

First-order Markov model probability in the coding region of mtb48 gene will be

$$\begin{aligned} P(X) &= P(A) * P(AG) * P(GT) * P(TA) * P(AG) * P(GC) \\ &\quad * P(CT) * P(TT) * P(TC) * (CC) * P(CA) * P(AG) \\ &= 0.204066 * 0.266904 * 0.119097 * 0.116162 \\ &\quad * 0.266904 * 0.326489 * 0.143902 * 0.151515 \\ &\quad * 0.323232 * 0.263415 * 0.20 * 0.266904 \\ P &= 6.50694\text{E}-09 \end{aligned}$$

**20.2.11. Determination
of the Most Likely Path
for Sequence
“CGCGTACTTCAATG”
in Frame 1 for the Zero
Order HMM Derived for
the mtb48 Gene in
Protocol 20.2.9.2,
Using the Following
Initial Transition
Probabilities**

$$\begin{aligned} a_{0c} &= a_{0n} = 0.5 \\ a_{nn} &= a_{cc} = 0.5 \\ a_{cn} &= 0.55 \\ a_{nc} &= 0.45 \end{aligned}$$

20.2.11.1. Hidden State Transition Probability

Hidden state Transition probability (provided in the protocol) can be summarized as follows:

		To			
From		A	T	G	C
	Coding	0.204	0.143	0.354	0.298
	Noncoding	0.195	0.186	0.334	0.285

20.2.11.2. Observable State Probability Matrix from the Zero-Order Markov Model for Coding and Noncoding Region of mtb48 Gene as Generated in Protocol 20.2.9.2 and 20.2.9.3

Coding	Noncoding
0.5	0.5

20.2.11.3. Starting Distribution (Provided in the Protocol)

	Coding	Noncoding
Coding	0.5	0.55
Noncoding	0.45	0.5

20.2.11.4. Determination of Most Likely Path Using Viterbi Algorithm

The most likely path for any of the sequence can be determined using Viterbi algorithm. The Viterbi algorithm is a computationally efficient method for determining the most apparent path taken through a Markov graph [8].

For the problem we have in our hand, the most likely path of the sequence CGCGTTACTTCAATG in frame 1 of the mtb48 gene can be calculated based on the HMM as shown Fig. 20.8.

The probability of the first base “C” in the sequence “CGCGTTACTTCAATG” to be in the coding/noncoding region can be calculated as

$$\begin{aligned}
 P_{\text{Coding}} &= P(C|\text{coding}) * \text{Start coding} \\
 &= 0.298 * 0.5 \\
 &= 1.49 * 10^{-1},
 \end{aligned}$$

while the probability of “C” to be in the noncoding region can be calculated as

$$\begin{aligned}
 P_{\text{Noncoding}} &= P(C|\text{noncoding}) * \text{Start noncoding} \\
 &= 0.285 * 0.5 \\
 &= 1.425 * 10^{-1}.
 \end{aligned}$$

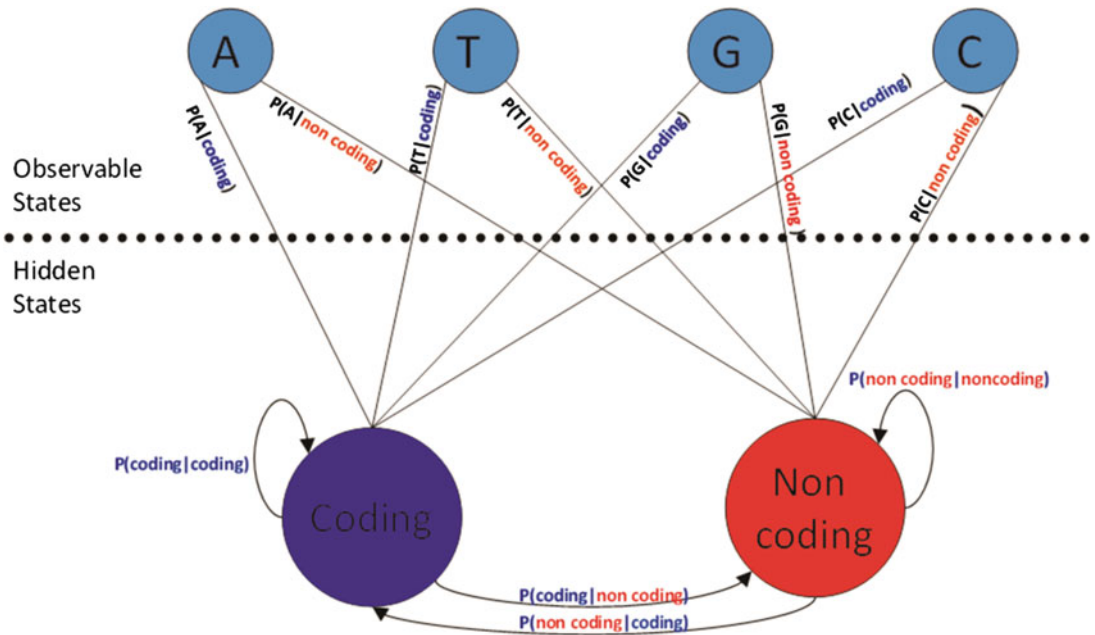


Fig. 20.8 Hidden Markov Model for the coding/noncoding region mapping.

From the above two probabilities of the first base “C” to be in the coding or noncoding region, the probability is higher for the coding region. Thus it is likely that the given sequence will start from the coding region.

For the second base, i.e., “G” in the sequence “CGCGTTACTTCAATG” to be in the coding and noncoding region, probability can be calculated as

$$\begin{aligned}
 P_{\text{Coding}} &= \max \left\{ \begin{array}{l} P(\text{coding}|\text{coding}) * P_{n-1}(\text{coding}), \\ P(\text{coding}|\text{noncoding}) * P_{n-1}(\text{noncoding}) \end{array} \right\} * P(G|\text{coding}) \\
 &= \max \{ (0.5 * 0.298 @ 0.45 * 0.285) \} * 0.354 \\
 &= 5.274 * 10^{-2},
 \end{aligned}$$

$$\begin{aligned}
 P_{\text{noncoding}} &= \max \left\{ \begin{array}{l} P(\text{noncoding}|\text{coding}) * P_{n-1}(\text{coding}) \\ P(\text{noncoding}|\text{noncoding}) * P_{n-1}(\text{noncoding}) \end{array} \right\} \\
 &\quad * P(G|\text{noncoding}) \\
 &= \max \{ (0.55 * 0.298 @ 0.5 * 0.285) \} * 0.334 \\
 &= 5.474 * 10^{-2}.
 \end{aligned}$$

From the above two probabilities of the second base “G” to be in the coding or noncoding region, the probability is higher for the noncoding region. Thus it is likely that the second nucleotide of the given sequence will come from the noncoding region. Similarly, the probability of all the bases in the sequence from coding/noncoding region can be calculated. The probability calculated for all the positions is given in Table 20.1. The larger

Table 20.1
The probability calculated for all the positions of the sequence CGCGTTACTTCAATG using Viterbi algorithm

	Coding	Noncoding
C	0.149	0.1425
G	0.052746	0.0547426
C	0.052746	0.0554895
G	0.052746	0.0554895
T	0.025311	0.0362142
T	0.0119691	0.017298
A	0.0170748	0.018135
C	0.030396	0.031977
T	0.021307	0.0304854
T	0.0119691	0.017298
C	0.0249426	0.026505
A	0.030396	0.0319605
A	0.020808	0.021879
T	0.014586	0.0208692
G	0.0296298	0.031062

probability value among the coding and noncoding regions is highlighted as the yellow background and shows the most likely path.

Thus, the most likely path of the given sequence can be identified as

CGCGTTACTTCAATG
c nc nc nc nc nc nc nc nc nc nc nc nc nc nc
c = coding and nc = noncoding

References

1. Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 48:443–453

2. Smith TF, Waterman MS (1981) Identification of common molecular subsequences. *J Mol Biol* 147:195–197

3. Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7:203–214

4. Feng DF, Doolittle RF (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J Mol Evol* 25:351–360

5. Bucka-Lassen K, Caprani O, Hein J (1999) Combining many multiple alignments in one improved alignment. *Bioinformatics* 15: 122–130
6. Bruno WJ, Socci ND, Halpern AL (2000) Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol Biol Evol* 17:189–197
7. Baum LE, Petrie T (1966) Statistical inference for probabilistic functions of finite state Markov chains. *Ann Math Stat* 37:1554–1563
8. Forney GD (1973) The Viterbi algorithm. *Proc IEEE* 61:268–278

Appendix 1: Biosafety, GLP, and Biosecurity

Each student and teachers who are a part of the laboratory should follow some guidelines strictly inside the laboratory for their own safety measure. The following guidelines have been provided to maintain a safe laboratory environment. It is the responsibility of each person who enters into the laboratory to understand the safety and health hazards associated with potential hazardous materials and equipment in the laboratory. It is also the individual's responsibility to practice the following general safety guidelines at all times:

General Safety Guidelines

1. Always wear proper eye protection in chemical work, handling, and storage areas. Contact lenses should normally not to be worn.
2. Always wear appropriate protective clothing with a suitable lab coat or apron.
3. Only closed toe shoes are to be worn in the laboratory. Sandals are not permitted.
4. Confine long hair and loose clothing.
5. Work areas or surfaces must be disinfected before and after use.
6. Always wash hands and arms with soap and water before leaving the work area. This applies even if you have been wearing gloves.
7. Never engage in horseplay, pranks, or other acts of mischief in biological work areas.
8. Label all materials with your name, date, and any other applicable information.
9. Do not pour biohazardous fluids down the sink.

10. Be familiar with the location of emergency equipment—fire alarm, fire extinguisher, emergency eyewash, and safety shower. Know the appropriate emergency response procedure.
11. Never mouth pipette chemicals when transferring solutions. Instead always use a pipet bulb to transfer solutions.
12. Report any accident however minor immediately.

Eating, Drinking, and Smoking

Eating, drinking, smoking, gum chewing, applying cosmetics, and taking medicines in laboratories are strictly prohibited.

1. Food, beverages, cups, and other drinking and eating utensils should not be stored in areas where hazardous materials are stored or handled.
2. Glassware used for laboratory operations should never be used to prepare or consume food or beverages.
3. Laboratory refrigerators, ice chests, cold rooms, ovens, and so forth should not be used for food storage or preparation.
4. Laboratory water sources and deionized water should not be used for drinking water.
5. Laboratory materials should never be consumed or tasted.

Housekeeping and Maintenance

In the laboratory keeping things clean and organized can help provide a safer environment. Keep drawers and cabinet doors closed and electrical cords off the floor to avoid tripping hazards. Keep aisles clear of obstacles such as boxes, chemical containers, and other storage items that might be put there. Avoid slipping hazards by cleaning up spilled liquids promptly and by keeping the floor free of loose equipment such as stirring rods, glass beads, stoppers, and other such hazards. Never block or even partially block the path to an exit or to safety equipment, such as a safety shower or fire extinguishers. Use the required procedure for the proper disposal of chemical wastes and solvents. Supplies and laboratory equipment on shelves should have sufficient clearance so that, in case of a fire, the fire sprinkler heads are able to carry out their function. The work area should be kept clean and uncluttered, with hazardous materials and equipment properly stored. Clean the work area upon completion of a task and at the end of the day. The custodial staff is only expected to perform routine duties such as cleaning the floor and emptying the general trash. In preparation for any maintenance service such as fumehood repair, plumbing, electrical, etc., the laboratory staff must prepare the laboratory before the maintenance personnel arrive. Whenever possible remove hazards that maintenance personnel may encounter during their work activities. For example, infectious agents, radioactive materials, or chemicals must be moved to a secure area prior to initiation of maintenance work.

**Biohazard Waste
Disposal****Dispose of items in the special receptacles as indicated below:**

Material	Method of disposal
Agar slants with biological materials	Place tube upright in indicated test tube rack, but place caps in baskets as indicated
Biological liquid (not in test tubes)	Leave in container with closed cap
Biological liquid in test tubes	Place tube upright in indicated test tube rack, but place caps in baskets as indicated
Broken glass (contaminated)	Sharps container
Broken glass (not contaminated)	Broken glass container
Cotton swabs (contaminated)	Benchtop disinfectant/discard can
Needles, glass slides, syringes, pipettes, other types of sharps	Sharps container
Noncontaminated paper	Regular trash
Petri dishes and contaminated solids (other than pipettes or swabs)	Biohazard “orange/red bag” container
Transfer pipettes (contaminated)	Benchtop disinfectant/discard can

Appendix 2: Address for Instruments and Chemicals Suppliers

Instrument Manufacturers

Bio-Rad (Asia Pacific) Pvt. Ltd, USA
Bio-Rad Laboratories (Singapore) Pvt. Ltd
27 International Business Park
#01-02 iQuest @ IBP
Singapore 609924, Singapore

PerkinElmer
940 Winter Street
Waltham, MA 02451, USA

Olympus
31 Gilby Road, Mount Waverely
VIC 3149, Australia
Tel: +61-3-9265-5400
Fax: +61-3-9562-6438
E-mail: Info@olympus.com.au

Nikon Corporation Ltd, Japan
5-21, Katsushima 1-chome
Shinagawa-ku, Tokyo 140-0012, Japan
Tel: +81-3-5762-8911

Labmate (Asia Pacific) Pvt. Ltd, UK
Labmate (Asia) Pvt. Ltd
Baid Mehta Complex, C-Block

183 Mount Road
Chennai 600015, India

UVP Ltd, UK

Ultra-Violet Products Ltd
Unit 1, Trinity Hall Farm Estate
Nuffield Road, Cambridge CB4 1TG, UK

Millipore SAS, France

Millipore S.A.S – Molsheim
BP 116
67124 Molsheim Cedex, France

Tomy Digital Biology Co., Ltd, Tokyo

3-14-17 Tagara, Nerima-ku
Tokyo 179-0073, Japan
E-mail: info@digital-biology.co.jp

Eppendorf AG., Germany

Barkhausenweg 1
22339 Hamburg, Germany

Shimadzu Asia Pacific Pvt. Ltd, Singapore

79 Science Park Drive
#02-01/08 Cintech IV
Singapore Science Park I
Singapore 118264, Singapore

Sonics & Materials Inc., USA

Sonics & Materials Inc.
53 Church Hill Road
Newtown, CT 06470-1614, USA
203.270.4600 – 800.745.1105 – 203.270.4610

Applied Biosystem Internationals, USA

Lingley House

120 Birchwood Boulevard
Warrington, Cheshire WA3 7QH, UK

Sartorius AG, Germany

Sartorius AG Weender Landstr.
94-108 37075 Goettingen, Germany

New Brunswick Scientific Co Inc., USA

Kerkenbos 11016546
BC Nijmegen, Netherlands
Tel: +31-24-3717-600

Castel MAC SpA, Italy

Via del Lavoro, 931033
Castelfranco Veneto, Province of Treviso, Italy
Tel: +39-0423738451

Astec Co. Ltd, Japan

Horiguchi Bldg., 2-19-7
Iwamoto-cho, Chiyoda-ku
Tokyo 101-0032, Japan

Molecular Device Corporation, Singapore

Molecular Devices, LLC
1311 Orleans Drive
Sunnyvale, CA 94089-1136, USA

Thermo Fisher Scientific (Asheville), LLC

308 Ridgefield Court Asheville
NC 28806, USA
Tel: +1-828-658-2711

Scigenics Biotech

35 (18), Vasudevapuram
Triplicane, Chennai 600 005, India

Chemical Suppliers

Fermentas Inc.

798 Cromwell Park Drive

Suites R-S

Glen Burnie, MD, USA

Tel: +1-800-3409026

Fax: +1-800-4728322

E-mail: fermentas.info@thermofisher.com

Sigma-Aldrich Company Ltd

The Old Brickyard

New Road

Gillingham

Dorset, SP8 4XT, UK

Roche Diagnostics Corp.

9115 Hague Road

Indianapolis, IN 46250, USA

Tel: +1-317-521-2000

Promega Corporation

2800 Woods Hollow Road

Madison, WI 53711, USA

Tel: +1-608-274-4330

Fax: +1-608-277-2516

Qiagen Inc.

27220 Turnberry Lane

Valencia, CA 91355, USA

Tel: +1-800-426-8157

Fax: +1-800-718-2056

Technical: 800-DNA-PREP

Tel: +1-800-362-7737

HiMedia Laboratories

A-406, Bhaveshwar Plaza, LBS Marg

Mumbai, Maharashtra 400086, India

Tel: +91-22-25003747/0970

BD

1 Becton Drive
Franklin Lakes, NJ 07417, USA
Tel: +1-201-847-6800

Difco International B.V.

Europaplein 30-S
8916 HH Leeuwarden
The Netherlands

Merck & Co., Inc. Global Headquarters

One Merck Drive
P.O. Box 100
Whitehouse Station, NJ 08889-0100, USA
Tel: +1-908-423-1000

Oxoid Limited

Wade Road
Basingstoke
Hampshire, RG24 8PW, UK

INDEX

A

- Actinomycetes DNA isolation
 - materials 7
 - method 11
- Affymetrix GeneChips 264
- Agarose/acrylamide, microbial DNA isolation
 - materials 8
 - method 14
- Agarose gel electrophoresis
 - materials 78–80
 - method 82–84
 - uses 75
- ALN/Clustal format 215–218
- Amplified fragment length polymorphism (AFLP) 40–41
 - amplification 49
 - banding 50
 - DNA digestion 49
 - fragment loading 49–50
 - procedure 50–52
 - requirements 44–45
- Amplified ribosomal DNA restriction analysis (ARDRA)
 - genetic fingerprinting 42, 61–62
 - methylophilic bacteria 130
 - molecular phylogenetics 251, 254
- Autotrophic methylophilic 124

B

- Bacillus* species, molecular identification
 - ARDRA and partial sequencing, 16S
 - rDNA 102–103
 - partial sequencing, 16S rDNA 101–102
 - 16S rDNA *vs.* 16–23S ITS nucleotide sequences 100–101
- Bioinformatics tools
 - materials
 - DNA and proteins, sequence analysis tools
 - for 218
 - nucleotide and protein sequence analysis 218–219
 - raw sequence editing tool 217
 - sequence format conversion, seq
 - Verter 216–217
 - methods
 - DNA and proteins, sequence analysis tools
 - for 222–226

- metabolomics and metagenomics, bioinformatics
 - tools for 241, 242
- multiple sequence alignment,
 - CLUSTALW 235–238
- nucleotide and protein sequence
 - analysis 225, 227–233
- nucleotide sequence similarity searching tool,
 - BLASTN 230, 232–235
- primer sets 239–241
- raw sequence editing tool 221–223
- sequence format conversion, seq
 - Verter 219–221
- sequence format
 - ALN/Clustal format 215–218
 - EMBL format 212–214
 - FASTA format 210
 - GenBank format 210–212
 - MSF file format 212, 215
 - PHYLP file format 212, 214–216
- Biological sequence analysis
 - DNA fragment probability 329
 - hidden Markov model
 - assumptions 322–323
 - emission probability 326
 - five sequence alignment 323–324
 - generation 322
 - matching, insert and delete 322
 - probability matrix calculation 324
 - profile parameters 324–325
 - transition probability 325–326
 - Markov model
 - first-order model 327–328
 - K-order model 326–327
 - zero-order model 327
 - most likely path
 - hidden state transition probability 330
 - viterbi algorithm 330–332
 - multiple sequence alignment
 - dynamic programming 310–311
 - Greedy approach, 312–313 (*see also* Greedy approach)
 - homologous protein sequences 309–310
 - pairwise alignments 311
 - uses 309
 - pairwise sequence alignment 303–304

- Biological sequence analysis (continued)
- construction and initialization,
 - of matrix 304, 305
 - global/local pairwise alignment 304
 - matrix filling 304–306
 - trace back, optimal alignment 306–309
 - phylogenetic tree construction
 - neighbor joining algorithm 318–319
 - UPGMA and neighbor joining
 - method 319–321
 - progressive alignment method
 - algorithm initialization 318
 - iteration 318
 - optimal pairwise alignment 316–317
 - phylogenetic tree construction 317
 - UPGMA algorithm 318
- Blunt end cloning 298
- Bromophenol blue 90
- C**
- cDNA fragmentation 283, 286
- Cell extraction, for microbial DNA isolation 4
- Competitive ELISA
- advantages and limitations 176
 - coating and blocking 182
 - incubation and detection 183
 - protocol of 174, 176
- Cryopreservation
- definition 146
 - materials 147
 - methods 147–149
- Cucumber mosaic virus (CMV) 18–19
- D**
- Denaturing gradient gel electrophoresis (DGGE)
- genetic fingerprinting 43
 - procedure 65–67
 - requirements 46
 - methylophilic bacteria 132
 - molecular phylogenetics
 - banding patterns and statistical
 - analysis 253–254
 - casting 253
 - gel preparation 253
 - loading buffer, preheating of 252
 - materials 250–251
 - PCR reaction 252
 - sample preparation 252
 - staining and documenting gel 253
 - uses 247
- Direct ELISA
- advantages 171
 - coating and blocking 180
- incubation and detection 180
- limitations 171
- protocol of 171, 172
- Directional cohesive end cloning 298
- DNA
- cloning and sequencing
 - competent cell preparation 295–296
 - excess salt and enzyme removal 300
 - foreign DNA fragment 291
 - Exo-SAP digestion 300, 301
 - experimental procedure 298
 - insert, release of 300
 - ligation 294–295, 299
 - materials 292
 - PCR product cleanup 302
 - PCR product cloning 298–299
 - PCR sequencing 300–302
 - phosphatase-treated vector preparation 294
 - positive clones selection 297–298
 - sequence analysis 302
 - transformation 295, 296
 - transformation and selection, clones 299
 - vector and foreign DNA fragment
 - preparation 293
 - concentration 5
 - extraction and purification (*see* Microbial DNA
 - extraction and quantitation)
 - fingerprinting/profiling/testing (*see* Genetic
 - fingerprinting)
 - microarrays 262, 264
- Drop dialysis method
- materials 8
 - method 13
- Dynamic programming (DP)
- global/local pairwise alignment 304
 - matrix construction and initialization 304, 305
 - matrix filling 304–306
 - progressive alignment method 315–316
 - steps in 304
 - trace back, optimal alignment 306–309
- E**
- Electrocompetent DH10B cells 206–207
- Electroelution 188
- Electrophoresis
- agarose gel
 - melting 78, 80
 - method 82–84
 - resolution of 78, 79
 - running buffers 78, 79
 - uses 75
 - apparatus 74
 - buffers 76–77
 - description 73

- DGGE (*see* Denaturing gradient gel electrophoresis (DGGE))
- limitations74
- SDS-PAGE (*see* Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE))
- sieving73
- TGGE (*see* Temperature gradient gel electrophoresis (TGGE))
- Electroporation method
- electro-competent DH10B cells
- preparation130–131
- ligation131
- transformation131–132
- EMBL format212–214
- Enzyme-linked immunosorbent assay (ELISA)
- competitive ELISA174–176
- direct ELISA171, 172
- indirect ELISA171–174
- materials
- coating buffer179
- reagents179–180
- methods180–183
- multiplex ELISA176–178
- principle170–171
- procedure169
- sandwich ELISA174, 175
- uses169
- Escherichia coli*
- chemical lysis2
- chromosomal DNA2
- genomic DNA isolation2–3
- materials5–6
- method9
- Ethidium bromide77
- Export single sequence files widget220
- Extremophiles, 153–154. *See also* Molecular identification procedures
- F**
- FASTA format210
- Fluorescence in situ hybridization
- materials155–156
- methods
- counterstain, DAPI160
- FISH probe-EUB388159
- hybridization and washing buffer
- preparation159, 160
- polycarbonate filters159–160
- sediment samples, fixation and
- preparation of158
- Fluorescence resonance energy transfer (FRET)26–27
- Fluorescent-based detection, of viral gene. *See* Viral gene quantitation, fluorescent-based
- Freezing
- description136, 138
- freeze-drying (*see* Lyophilization)
- glycerol139
- Fungal DNA isolation, CTAB method
- materials8
- method11
- G**
- Gel electrophoresis
- Macrophomina phaseolina*97
- for microbial DNA isolation
- materials9
- method94–15
- GenBank format210–212
- Genetic fingerprinting
- amplified fragment length polymorphism40–41
- amplification49
- banding50
- DNA digestion49
- fragment loading49–50
- procedure50–52
- requirements44–45
- amplified ribosomal DNA restriction analysis42, 61–62
- denaturing/temperature gradient gel
- electrophoresis43
- procedure65–67
- requirements46
- genes and primer sequences39
- hybridisation method38
- PCR method38, 70
- random amplified polymorphic DNA40
- annealing47–48
- banding pattern47
- markers generation46
- primer44
- products separation47–48
- requirements43–44
- RFLP43
- procedure67–70
- requirements46
- ribosomal RNA genes38
- simple sequence repeats/short tandem repeats41
- banding54
- hybridisation53
- microsatellite loci identification53
- procedure55–58
- requirements45
- single strand conformation polymorphism41–42
- amplification products60
- banding59
- procedure60–61
- requirements45

- Genetic fingerprinting (continued)
- terminal restriction fragment length
 - polymorphism 42–43
 - amplification 62
 - fluorescently labelled primers 62, 63
 - procedure 64–65
 - requirements 45–46
- Gram negative bacterial DNA isolation, CTAB
- method 6
 - materials 10
- Greedy approach 312–313
- designing pairs and identification
- of alignment 314
 - multiple sequence alignment 315
 - new profile creation 314–315
 - pairwise alignment performance 313–314
 - progressive alignment method 315–316
 - two sequence combination 314
- GyrA* gene sequencing 252, 256–257

H

- Hidden Markov model (HMM)
- assumptions 322–323
 - emission probability 326
 - five sequence alignment 323–324
 - generation 322
 - matching, insert and delete 322
 - probability matrix calculation 324
 - profile parameters 324–325
 - transition probability 325–326
- HindII* and *HindIII*, 31
- Hydrolysis probe 25–26

I

- Indirect ELISA
- advantages 172
 - blocking 181
 - coating 180–181
 - incubation and detection 181
 - limitations 174
 - protocol of 172, 173

L

- Laemmli method 76
- Laser scanning, microarray 275–276
- Lauryl sulfate 75
- Liquid paraffin, microbe preservation 137–138
- Lyophilization
- advantages 139
 - cultivation of bacteria 140–141
 - freezing 141
 - materials 143

- medium selection 140
- methods using
 - manifold 145
 - shelf lyophilizer 143–145
- post-lyophilization storage 143
- primary drying 142
- secondary drying 142

M

- Macrophomina phaseolina*, molecular identification
- distribution 93
 - gel electrophoresis 97
 - isolation 94–95
 - materials 94
 - microscopy 95
 - PCR 95–96
- Markov model (MM)
- first-order model 327–328
 - K-order model 326–327
 - zero-order model 327
- Maxam-Gilbert method 76
- Maximum likelihood method 250
- Metabolomics analysis tools 241, 242
- Metagenomic libraries
- construction of 192, 200–207
- DNA isolation
- CTAB method 190, 193
 - indirect lysis method 191, 194
 - TNEP method 190–191, 193–194
- DNA purification
- agarose gel 191, 195
 - Sephadex G-50 beads 191, 194–195
- RNA isolation 191–192
- 16S rRNA amplicon cloning 192, 197–198
- 16S rRNA sequences 192, 196–197
- phylogenetic analysis 198–200
- Metagenomics analysis tools 241, 242
- Methylophilic bacteria, molecular identification of
- amplified ribosomal DNA restriction analysis 130
 - BLAST search 133
 - cloning 130–132
 - cultivable and uncultivable 127
 - DGGE 132
 - genomic and community DNA extraction 129
 - materials 128
 - PCR 125
 - phylogenetic analysis 133
 - primers 126
 - real time PCR 133
 - RFLP 130
 - 16S rRNA and *mxoA* gene
 - amplification 129
 - sequencing 132–133
 - traditional identification method 124–125

- Microarrays
- affymetrix GeneChips264
 - data normalization and statistical analysis ..277–278
 - disease diagnosis278
 - DNA microarrays262, 264
 - drug discovery278
 - experimental steps265–267
 - gene discovery278
 - materials283–284
 - method
 - cDNA fragmentation286
 - cDNA synthesis284–285
 - poly-A RNA control preparation284, 285
 - purification and quantification, cDNA286
 - RNA isolation284
 - RNA removal285
 - scanning289–290
 - stain reagents288
 - target hybridization286–288
 - terminal labeling286, 287
 - washing and staining288–289
 - microbiology278–279
 - principle262–263
 - probes281–282
 - protocols
 - fabrication267–268
 - hybridization272–275
 - laser scanning275–276
 - printing270–271
 - procedure268–269
 - sample labeling271–272
 - slide blocking269–270
 - Pseudomonas*, 282
 - spot quantization276–277
 - Staphylococcus*, 282
 - toxicological research278
 - types of263–264
 - uses261
- Microbial cultures
- continuous growth methods135–136
 - cryopreservation146–149
 - drying136
 - freezing
 - description136, 138
 - glycerol139
 - lyophilization139–145
 - mineral oil/liquid paraffin storage
 - materials137
 - methods138
 - periodical transfer to fresh media136
 - refrigeration137
 - safety
 - cryogenic storage149–150
 - culture handling149
 - freeze-drying150
- Microbial DNA extraction and quantitation
- actinomycetes
 - materials7
 - method11
 - agarose/acrylamide
 - materials8
 - method14
 - calculation method16
 - cell extraction4
 - DNA concentration5
 - drop dialysis method
 - materials8
 - method13
 - Escherichia coli*
 - chemical lysis2
 - chromosomal DNA2
 - genomic DNA isolation2–3
 - materials5–6
 - method9
 - fungal DNA isolation, CTAB method
 - materials8
 - method11
 - gel electrophoresis
 - materials9
 - method94–15
 - Gram negative bacterial DNA isolation, CTAB
 - method
 - materials6
 - method10
 - organic extraction4
 - phenol extraction and ethanol precipitation
 - materials8
 - method12–13
 - plasmid DNA
 - materials6–7
 - method10–11
 - molecular structure3–4
 - sephacryl S-500 spin columns
 - materials8
 - method13
 - spectrophotometer
 - materials9
 - method15–16
 - yeast
 - materials7
 - method11
- Microsatellites41
- Mineral oil, microbe preservation137–138
- Mini-slab gel system84
- Molecular beacons27, 28
- Molecular identification
- Bacillus* species
 - ARDRA and partial sequencing, 16S
 - rDNA102–103
 - partial sequencing, 16S rDNA101–102

Molecular identification (continued)

- 16S rDNA *vs.* 16-23S ITS nucleotide sequences 100–101

Macrophomina phaseolina

- distribution 93
gel electrophoresis 97
isolation 94–95
materials 94
microscopy 95
PCR 95–96

materials

- fluorescence in situ hybridization 155–156
PCR-denaturing gradient gel electrophoresis 156–157
16S rDNA sequencing 158
16S rRNA library construction 158
terminal restriction fragment length polymorphism 157–158

methods

- fluorescence in situ hybridization 158–160
PCR-denaturing gradient gel electrophoresis 161–163
16S rDNA sequencing 165–167
16S rRNA library construction 165–167
terminal restriction fragment length polymorphism 163–165

methylophilic bacteria (*see* Methylophilic bacteria, molecular identification of)*Pseudomonas*

- PCR 106, 108, 109
16S rRNA gene amplification 107–109
16S rRNA-RFLP 107, 109–110

Vibrios

- flow chart for identification 114
full-length *toxR* gene analysis 116
infection 115
materials 116–118
methods 118–122
multilocus sequence analysis 115
multiplex PCR 115–116
pulse field gel electrophoresis 115
serology testing 115

Molecular phylogenetics

- amplified rDNA restriction analysis 247
genetic fingerprinting technique 246
gyrA 248–249
materials
ARDRA 251
DGGE 250–251
gyrA gene sequencing 252
rpoB gene sequencing 251
16S rRNA gene sequencing 251
maximum likelihood method 250

methods

- ARDRA 254
cladogram 259
clustal X 257–259
DGGE 252–254
gyrA gene sequencing 256–257
rpoB gene sequencing 255–256
16S rRNA gene sequencing 255
MSA 249
parsimony method 250
RFLP 248
rpoB 248
SSU rDNAs 246

Molecular taxonomy. *See* Molecular phylogenetics

Multilocus sequence alignment (MSA) 249

Multiple sequence files (MSF) format 212, 215

Multiplex ELISA

- advantages 177–178
analysis of data 183
applications 178
coating and blocking 183
detection and incubation 183
limitations 178
protocol 177

O

Oligo microarray 263

Organic extraction 4

P

Parsimony method 250

PCR-denaturing gradient gel electrophoresis

- materials 156–157
methods
DGGE conditions 161
DNA extraction 161–162
phylogenetic analyses 163
purified PCR products and cloning, sequencing of 162

Pearson format. *See* FASTA format

Phenol extraction and ethanol precipitation

- materials 8
method 12–13

Phylogenetic analysis tools 238

Phylogeny 126

Phylogeny Inference Package (PHYLIP) file

- format 212, 214–216

Plasmid DNA isolation

- materials 6–7
method 10–11
molecular structure 3–4

Potato dextrose agar (PDA) 136

Primer/probe designing tools	240
Printing, in microarray technology	270–271
Pseudomethylotrophs	124
<i>Pseudomonas</i> , molecular identification	
PCR	106, 108, 109
16S rRNA gene amplification	107–109
16S rRNA-RFLP	107, 109–110

R

Random amplified polymorphic DNA (RAPD)	40
annealing	47–48
banding pattern	47
markers generation	46
primer	44
products separation	47–48
Raw sequence editing tool	217, 221–223
Refrigeration	137
Restriction digestion, metagenomic DNA	200–201
Restriction enzymes	
coding sequences	32–33
fragment length study	32
materials	35
method	35–36
repetitive DNA	33
role of	31–32
specificity recognition	33
star activity	33–34
sticky/cohesive ends	32
termini	34
types	32
Restriction fragment length polymorphism (RFLP)	248
<i>RpoB</i> gene sequencing	251, 255–256

S

Sandwich ELISA	
advantages	174
blocking and detection	182
coating	181
incubation, with HRP-conjugated antibody	182
limitations	174
protocol of	174, 175
standard and samples incubation	182
Sanger method	76
Scorpion probes	27–29
Sephacryl S-500 spin columns	
materials	8
method	13
SeqVerter tools	216–217, 219–221
Serological typing	170
Serotyping scheme	115
Sieving	73

Significance analysis of microarrays (SAM)	277
Silver staining SDS PAGE gels	
materials	81–82
method	90–91
Simple sequence repeats (SSRs)/short tandem repeats (STR)	41
banding	54
hybridisation	53
microsatellite loci identification	53
procedure	55–58
requirements	45
Single strand conformation polymorphism (SSCP)	41–42
amplification products	60
banding	59
procedure	60–61
requirements	45
Slide blocking	269–270
Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)	
Laemmli method	76
materials	
acrylamide	80, 81
silver staining gels	81–82
stacking gel solution	79
method	
cassettes	84
disassembly and staining	89
gel separation	84–86
loading	88–89
protein samples	87
sample denaturation	87–88
SDS gel preparation	84
stacking gel preparation	86
staining protein gels	89–90
Spectrophotometer, microbial DNA isolation	
materials	9
method	15–16
Spot quantization	276–277
<i>Staphylococcus</i> , 282	
T	
TA cloning	298
Temperature gradient gel electrophoresis (TGGE)	43
procedure	65–67
requirements	46
Terminal restriction fragment length polymorphism (T-RFLP)	
genetic fingerprinting	42–43
amplification	62
fluorescently labelled primers	62, 63
procedure	64–65
requirements	45–46

Terminal restriction fragment length polymorphism (T-RFLP) (continued)	
molecular identification procedures	
community DNA, isolation of	163–164
community similarity	164–165
materials	157–158
PCR amplification	164
species richness and community signature	164–165
T-RFLP analysis	164
Type II restriction enzymes	32

V

<i>Vibrios</i> , molecular identification	
flow chart for identification	114
full-length <i>toxR</i> gene analysis	116
infection	115
materials	116–118
methods	118–122
multilocus sequence analysis	115
multiplex PCR	115–116
pulse field gel electrophoresis	115
serology testing	115
Viral gene quantitation, fluorescent-based	
comparative threshold method	22–23
components	
primers and probe	20
probe selection criteria	20–21
template	21

cycling parameters	20
hybridization probes	
FRET	26–27
molecular beacons	27, 28
Scorpions	27–29
hydrolysis probe	25–26
instrument and set up	19
QRT-PCR reaction requirements	19
reaction set and thermal cycling	23
reagents	21
standard curve method	21–22
SYBER Green	24–25

W

Web interface, nucleotide BLAST search	
tool	220

X

Xylene cyanol	90
---------------	----

Y

Yeast DNA isolation	
materials	7
method	11